Exhibit B

REBUTTAL EXPERT REPORT OF PETER S. ARCIDIACONO

Students for Fair Admissions, Inc. v. University of North Carolina
No. 14-cv-954 (M.D.N.C.)

TABLE OF CONTENTS

1. Executive Summary

In my opening report, I concluded that non-URM applicants to UNC have significantly stronger qualifications than URM applicants; that UNC's admissions process is guided by an implicit formula; and that race plays a dominant role in individual admissions decisions (especially for out-of-state applicants).

I illustrated the magnitude of UNC's racial preferences with an example of a hypothetical male, non-FGC Asian-American applicant whose observed characteristics would imply a 25% chance of admission. If he were an in-state applicant, his probability of admission would increase to more than 63% (i.e., more than double) had he been treated like a Hispanic applicant, and to more than 88% (more than triple) had he been treated like an African-American applicant. If he were an out-of-state applicant, his probability of admission would increase to more than 85% had he been treated like a Hispanic applicant, and to more than 99% had he been treated like an African-American applicant. Simply changing this hypothetical Asian-American applicant's race to either Hispanic or African-American (with all other characteristics remaining the same) thus would transform him from an unlikely admit to an almost certain admit.

In addition, I concluded that holding fixed the number of admission slots and removing racial preferences would significantly change Asian-American and white representation at UNC. Fixing the number of in-state admits at the levels observed in the data and removing racial preferences would result in 1,219 additional non-URM admits over the six-year period, a 5.5% increase. Fixing the number of out-of-state admits at the levels observed in the data and removing racial preferences would result in 2,482 more non-URM admits from that pool, a 25.7% increase.

Professor Caroline Hoxby, who was hired by UNC to opine on these issues, submitted a competing report. In relevant part, Professor Hoxby

opines that "[e]mpirical analysis establishes that UNC admissions decisions cannot be explained using a formula containing verifiable student characteristics." In her view, then, UNC's admissions decisions "are consistent with a holistic review of candidates." She further opines that an "applicant's race does not determine UNC admissions decisions in a common and systematic way and it is not a dominant factor in admissions." (Hoxby at para. 6; pages 3-4.)

But Professor Hoxby makes several fundamental errors that undermine her analysis of UNC's admissions process. And though she claims an inability to model UNC admissions to produce accurate predictions, her model does just that. In doing so, her model actually confirms the existence of substantial racial preferences in favor of URM applicants.

Professor Hoxby begins by misunderstanding the relevant analytical framework. In response to my work demonstrating that race is a dominant factor in admissions decisions, Professor Hoxby states that this can be true only if "there exists a statistical formula of admissions decisions that accurately predicts outcomes and in which the race factor plays a dominant role." Naturally, she opines that UNC's admissions decisions are not formulaic and that race does not play a dominant role in those decisions. But she misapprehends and conflates two issues here. First, race may have an outsize role in admissions decisions regardless of whether UNC employs it via a formulaic or holistic approach. Second, formulaic admissions decisions do not rule out the use of subjective factors (for example, UNC's summary ratings) as part of the formula. Indeed, as to the presence of a formula (either an explicit or implicit one), the real question is whether race itself is considered in a formulaic manner; it makes no difference whether other factors are considered in a holistic manner.

On top of these analytical mistakes, Professor Hoxby errs by failing to account for all relevant factors in evaluating the role that race plays in

admissions decisions. In particular, Professor Hoxby excludes from her model the five summary ratings that UNC readers assign to applicants in evaluating their applications: (1) program, (2) performance, (3) extracurricular, (4) essay, and (5) personal quality. But, as I showed in my opening report, non-URM applicants are generally stronger than URM applicants with respect to these summary ratings. In failing to control for these ratings in her model, Professor Hoxby improperly attributes the effect of these ratings to race, thereby understating (and obscuring) the magnitude of UNC's racial preferences.

Professor Hoxby also errs by misinterpreting the fit of her model. That is, even setting aside her improper analytical framework, Professor Hoxby misinterprets her findings. Her key error is that she confuses the Pseudo-R square measure with the R-square—two fundamentally different metrics that relate to assessing the fit of a particular model. Properly accounting for the fit of her model shows that it—despite its other flaws—fits the data quite well.

What's more, Professor Hoxby's model actually demonstrates that UNC's racial preferences play an outsize role in the admission of URM applicants. In particular, her model indicates that *more than 40% of Hispanic admissions and more than 50% of African-American admissions are the result of racial preferences.*

Professor Hoxby makes several other errors and misguided coding choices that undermine her analysis. The bottom line is that nothing in her analysis takes away from my initial conclusions that UNC's admissions process is guided by an implicit formula in which race plays a dominant factor in individual admissions decisions.

2. Professor Hoxby makes several fundamental errors that undermine her analysis of UNC's admissions process.

2.1 Professor Hoxby misunderstands the relevant analytical framework.

In response to SFFA's claims that race is a dominant factor in admissions decisions, Professor Hoxby makes the following argument:

> If Plaintiff is correct, then the data will show that admissions decisions are formulaic and predicted by race. From a statistical perspective, this means that, if Plaintiff is correct, there exists a statistical formula of admissions decisions that accurately predicts outcomes and in which the race factor plays a dominant role. If no such formula can be derived, then that establishes that race is not a dominant factor and that the process is holistic. (para. 29, page 13)

Motivated by this argument, Professor Hoxby outlines stringent criteria for establishing whether UNC's admissions are formulaic. In her view, admissions must be functions of "verifiable measures" (points 37-38, page 15). These "verifiable measures" must be things taken directly from the application itself, with no room for judgment by the reader of the application. Professor Hoxby argues that this rules out using any of UNC's summary ratings (Program, Performance, Activities, Essays, and Personal Qualities) (point 39, page 16).

Professor Hoxby's arguments are incorrect for a number of reasons. First, race may have an outsize role on admissions decisions regardless of whether UNC employs it via a formulaic or holistic approach. Second, formulaic admissions do not rule out the use of subjective factors (for example, UNC's summary ratings) as part of the formula. Third, in order to assess the role of race in admissions one must account for *all* the relevant factors.

2.1.1 Race can be a dominant factor in admissions even under holistic admissions.

Professor Hoxby understands correctly that governmental use of race as part of a formula for making admissions decisions could present legal problems. But she is wrong to suggest that if UNC's admissions calculus is

"holistic," then race cannot be a dominant factor in UNC's admissions decisions.

Implicit in a holistic admissions decisions model is that, even with all the observed measures, there is uncertainty in the admissions decision. But this does not preclude race from being a dominant factor in the decision. For example, suppose that for a particular racial group the preferences given for race were much more important to their chances of admission than the "unobserved factors," that is, the factors that made the admissions decisions holistic in the first place. And suppose that these preferences were so large that the average admissions probability for this racial group more than doubled. Surely one would see this as a dominant factor in admissions for this group of students. In fact, as I describe in Section 2.3, this is what Professor Hoxby's own additive model shows: African-American admit probabilities are more than twice as high as a result of racial preferences.

2.1.2 Admissions may be formulaic with respect to race even if subjective factors are included in the model.

Notwithstanding that her models fail to include many of the key factors in UNC's admissions decisions, Professor Hoxby's models show very strong racial preferences. Professor Hoxby contends that because UNC's five summary ratings are not themselves formulaic, they should be excluded from any analysis of whether UNC's admissions process is formulaic. Hoxby para. 39. As Professor Hoxby puts it, "even if I could add up the summary ratings and predict the admissions decision accurately (which, in fact, I cannot),[1] I would not have shown that the admissions decision was formulaic … A decision based on ratings that cannot themselves be explained by formulas is not a formulaic decision." Id.

[1] As explained below, Professor Hoxby's models actually predict admissions with a relatively high degree of accuracy. See section 2.2.2.

Excluding these measures because some reader judgment is involved is incorrect for three reasons. First, many of UNC's ratings *are* formulaic (or at least as formulaic as the assignment of grades by high school teachers—and Professor Hoxby correctly includes high school grades/GPA in her model). For example, UNC's program rating is essentially a count of the number of advanced courses applicants have taken. Rosenberg Depo. 243:20-23. It is more objective and formulaic than a grade in a high school English class that requires evaluating the quality of a paper and similarly more objective and formulaic than the score given on the writing portion of the SAT or ACT (both of which Professor Hoxby deems appropriate for inclusion in her analysis).

Second, it does not matter whether there are subjective aspects to UNC's summary ratings if UNC's admissions decisions are formulaic *with respect to race*. To illustrate the concept, consider an art competition where paintings are scored on various dimensions (necessarily subjective) with the highest scores receiving prizes. Suppose that there was some concern that members of a certain race/ethnicity were underrepresented with respect to receiving awards for their artwork and, to boost their representation among award winners, some additional number of points were added to each of the scores for members of the preferred race/ethnicity. Professor Hoxby would not consider this formulaic, notwithstanding the fact that race would be applied in a formulaic manner.

Finally, the point system that the Supreme Court found unlawful in *Gratz v. Bollinger* **included** subjective factors. The various factors on the University of Michigan's scoring sheet required subjective judgments that were then fed in to a point system: a school factor, a curriculum factor, outstanding essay, personal achievement at different levels, and leadership achievement at different levels. Notwithstanding that those factors required subjective judgments, the University of Michigan's admissions decisions were quite obviously formulaic. But according to her criteria, Professor Hoxby would conclude otherwise.

2.1.3 It is important to account for all the relevant factors when determining the role of race in admissions.

In scoring applicants, UNC gives them five distinct "summary" ratings (as Professor Hoxby calls them), one for each of five different categories: (1) program, (2) performance, (3) extracurricular, (4) essay, and (5) personal quality. See Rosenberg Depo. 142:18-144:12; see also Hoxby para. 20. Professor Hoxby takes the surprising position that one should not control for the summary ratings when using statistical methods to analyze the effect of race on admissions decisions. Not controlling for these ratings results in omitted variable bias. Omitted variable bias is a condition where a statistical model attributes the effect of missing variables to the estimated effect of the included variables. This bias can affect both the magnitude and the sign of the estimated effects of the included variables. Returning to the art competition example, not accounting for the (necessarily subjective) scoring of the paintings would result in a substantial underestimate of the role of race in the awarding of prizes.

In other words, when Professor Hoxby models the UNC admissions process to determine the effect of race, her model (which does not include the five summary ratings) will improperly attribute the effect of those ratings to race. To help illustrate the effect of this omitted variable bias, take the following example: suppose that UNC values students who take more advanced classes (as indicated by their program rating).[2] Suppose further that group A took more advanced classes than group B. Finally, suppose further that UNC gives a preference for members of group B because of their race. Not accounting for the number of advanced classes applicants take will likely understate the extent of racial preferences for group B over group A as

[2] Rosenberg Depo. 243:17-244:23.

the estimated racial preferences will in part reflect differences in the number of advanced classes each group has taken.

This is precisely what my analysis reveals: because URM applicants are weaker on UNC's program, performance, extracurricular, and essay ratings than non-URM applicants, [3] not controlling for these ratings understates the magnitude of racial preferences that UNC gives to URM applicants. Properly controlling for UNC's summary ratings (as I do) thus reveals that UNC's racial preferences are much larger than Professor Hoxby's model indicates.[4] In other words, Professor Hoxby's assessment of the effect of race in the admission process, notwithstanding the other methodological problems with her approach, underestimates the magnitude of UNC's racial preferences because of her failure to control for UNC's summary ratings.

2.2 Professor Hoxby misinterprets the fit of her model.

The previous section outlined why I believe Professor Hoxby's framework for considering the role of race in admissions is flawed. But even under Professor Hoxby's framework, Professor Hoxby is incorrect in how she interprets her findings. The most fundamental mistake that Professor Hoxby makes concerns how she interprets the fit of her model. The key passage from Professor Hoxby's report reads:

---

[3] Arcidiacono Report 2, 25-26. URMs score higher on the personal quality rating, but my analysis indicates that this is a result of racial preferences in the rankings themselves. Arcidiacono 63-64.

[4] Note that to the extent UNC's racial preferences influence the summary ratings themselves, controlling for them will tend to underestimate the effect of race on UNC admissions decisions. There is evidence that UNC gives URM students a preference with respect to the personal quality rating, Arcidiacono Report 11-12, 56-57; Rosenberg Depo. 250:13-251:8; Perkins Depo. 40:3-43:10, which means that controlling for that rating will yield conservative estimates for the effect of race on admissions decisions, Arcidiacono Report 57, 64.

> The most widely accepted way to summarize whether a regression model (or formula based on one) explains a decision such as admit/reject is "R-squared." R-squared is a statistical measure that indicates how well the factors included in the regression explain the outcome. Roughly speaking, it is the percentage share of the admit/reject decision that a formula can predict. For instance, if the regression discovered a formula with an R-squared of 1.00, I could use the formula to predict the admissions decision and the prediction would be correct 100 percent of the time. However, if the regression discovered a formula with an R-squared of only 0.50, predictions based on that formula would be correct only 50 percent of the time. (para. 44 pages 18-19).

As I show below, this statement is wrong and misleading on multiple levels. Moreover, properly accounting for the fit of the model shows that her model—despite its other flaws—fits the data quite well.

2.2.1 Professor Hoxby confuses the Pseudo-R-square measure with the R-square, when they are fundamentally different metrics.

Professor Hoxby purports to use the R-square but actually uses the *Pseudo* R-square and makes a significant error in confusing the two. As a fundamental matter, the R-square is a measure of fit for models where the relationship between the error term is additive, while the Pseudo R-square is a measure of fit for models when the error is non-additive. Hence the Pseudo R-square is used for *non-linear* models, the R-square is used for *linear* models. The R-square measure applies to models where the estimated relationship between the outcome variable (in this case, admission/rejection) and the controls (e.g., test scores) is linear with an additive error term that represents the unobserved factors. Professor Hoxby would be correct that the R-square gives the share of the outcome variable that can be explained by the controls if the admissions model she estimated was linear; however, the model she estimates is *not* linear. To be sure, the *formula* underlying her

admissions model is linear, but the admissions model produced therefrom is *non*-linear.[5]

Again, it is *not* the R-square that Professor Hoxby shows in her report (as she claims), but the *Pseudo R-square*. Despite the similarity in names, the Pseudo R-square is a different measure from the R-square, and the two are in no way interchangeable. As measures, the only similarities the two have is that they run on a scale from 0 to 1, and higher values on either are indicative of a better fit of the data.

Most relevant here, Professor Hoxby's confusion regarding the two metrics causes her to suggest that her model is not a good fit—when in truth it is. The classic citation for what is considered an "excellent fit" based on the Pseudo R-square is McFadden (1979) page 307:

> Those unfamiliar with the ρ^2 index should be forewarned that its values tend to be considerably lower than those of the R^2 index and should not be judged by the standards for a 'good fit' in ordinary regression analysis. For example, values of 0.2 to 0.4 for ρ^2 represent an excellent fit.[6]

The ρ^2 referred to above later became known as McFadden's R-Square, or the Pseudo R-square. Note that once minimal controls are included, Professor Hoxby's models have Pseudo R-squares that are either in the 0.2 to 0.4 range or higher.[7] As I discuss in Section 2.2.2, there are other measures that can be used to evaluate the fit of models where the outcome is binary and these

---

[5] The two most common statistical methods for estimating models where the outcome is binary (as is the case with whether or not an applicant is admitted) are logit (which is what I use) and probit (which Professor Hoxby uses). Both of these are *non-linear* models where applicants are admitted when a *linear* combination of their observed and unobserved characteristics exceeds some threshold.

[6] D. McFadden, "Quantitative Methods for Analysing Travel Behavior: Some Recent Developments," Chapter 13 in Behavioral Travel Modeling, D.A. Hensher and P.R. Stopher, editors, Croom Helm Ltd., 1979.

[7] And once appropriate modifications are made to Professor Hoxby's model, the Pseudo R-squares are substantially higher than this range.

measures show Professor Hoxby's models provide substantial predictive power.

That the Pseudo R-Square and R-Square are different measures with different properties is well known and discussed in many econometrics textbooks. Christopher Dougherty, in his introductory econometrics textbook for undergraduates, explains the Pseudo R-squared thusly:[8]

> The pseudo-R^2 seen in some regression output, including that of Stata, compares its log-likelihood *log L*, with the log-likelihood that would have been obtained with only the intercept in the regression, $log\ L_0$. The pseudo-R^2 is the proportion by which *log L* is smaller, in absolute size, than $log\ L_0$ … While it has a minimum value of 0, its maximum value must be less than 1 and unlike R^2 it does not have a natural interpretation.[9]

The Pseudo R-square also became known as the likelihood ratio index, perhaps to avoid the mistake of confusing it with the true R-square. Indeed, in the classic econometrics book on the modeling of discrete outcomes Ken Train states:[10]

> It is important to note that the likelihood ratio index is not at all similar in its interpretation to the R^2 used in regression, despite both statistics having the same range. R^2 indicates the percentage of the variation in the dependent variable that is 'explained' by the estimated model.

---

8 Page 386, Christopher Dougherty, *Introduction to Econometrics, Fourth Edition*, 2011. Note that Stata is the program both Professor Hoxby and I use in our analysis of admissions.

9 As noted above, McFadden's work on Pseudo-R squares has yielded his well-accepted rule of thumb that Pseudo-R square values in the range of 0.2 to 0.4 represent an excellent fit.

10 Page 68, Kenneth E. Train, Discrete Choice Methods with Simulation (2d ed. 2009).

2.2.2 Professor Hoxby significantly understates how well her model fits the data.

Professor Hoxby's error here is thus a very fundamental one. But she compounds the error by contending "if the regression discovered a formula with an R-squared of only 0.50, predictions based on that formula would be correct only 50 percent of the time." Hoxby para. 44 page 19. This statement is not only incorrect; it would strike most experienced econometricians as nonsensical. To see why, it is helpful to consider a simple example. Suppose that a school admitted exactly half of its applicants, and suppose that our "model" predicting admissions relied wholly on some factor that was completely unrelated to admissions—for example, the flip of a coin. Our model would essentially be making random guesses about each applicant's chances, and because it had no true explanatory power, it would have a pseudo R-squared value of 0, or very close to 0. Yet this model would correctly "predict" admissions outcomes half of the time, since any random guess under these facts would have a 50% chance of being correct.[11]

Put simply, Professor Hoxby's use of the Pseudo R-squared is not appropriate for determining the extent to which UNC's admissions decisions are formulaic or how well her model correctly predicts admission. There are, however, alternatives. Namely, if the admission process was formulaic with the estimated coefficients giving the weights of the formula, then using the formula to predict admissions would generally get the predictions correct.

To illustrate, consider Professor Hoxby's richest additive model (model 9). For each of the admissions cycles Professor Hoxby uses, I can use the estimates of her model coupled with the observed characteristics of the applicants to create an *admissions index*. The admissions index provides a numerical summary of how strong each applicant is based on the applicant's observed characteristics (or, equivalently, the portion of the admissions

---

11 Indeed, the only way to get zero correct predictions would be to assign all the rejects as admits and all the admits as rejects.

formula that is due to observed characteristics). For each of the four admissions cycles Professor Hoxby considers, I rank the applicants according to their admissions index. Using this ranking, I then assign those with the highest rankings as those who would be predicted to be admitted based on the observables alone such that the number of admitted students matches what is observed for that admissions cycle. For example, if 1000 applicants are admitted in a particular cycle, then the 1000 applicants with the highest admissions index would be predicted to be admitted.

Table 2.1: Accuracy of Professor Hoxby's additive model

	Accuracy for Admits	Accuracy for Rejects	Overall Accuracy
Hoxby Additive Model	72.3%	89.9%	85.2%
Random Assignment	26.7%	73.3%	60.8%

Note: Hoxby Model refers to her additive model number 9. Accuracy is defined as the number of correct predictions divided by the appropriate sample size. For example, accuracy for admits is defined as the number of admits who were predicted to be admitted divided by the number of actual admits; overall accuracy is the number of admits who were predicted to be admitted plus the number of rejects predicted to be rejected divided by the number of applicants.

The first row of Table 2.1 shows how well Professor Hoxby's additive model (model 9) predicts who is admitted according to part of the formula associated with observed characteristics. Over 72% of those who are actually admitted would be predicted to be admitted based on their observables alone. And almost 90% of those who are actually rejected would be predicted to be rejected. As shown in the third column, the overall accuracy of the model—the total number of correct predictions divided by the number of applicants—is over 85%.

These percentages show that Professor Hoxby's model—despite its shortcomings—is actually a good predictor of UNC's admissions decisions. Furthermore, these percentages become more meaningful when compared to a random-assignment model—that is, a model that determines admission and

rejection by random assignment, again in such a way that the total number assigned to admission matches what is seen in the data. In this case, the overall admission (rejection) rate would give the fraction of the cases the model correctly predicted whether the applicant was actually admitted (rejected). This is what is presented in the second row of Table 2.1. Without any controls, less than 27% of the actual admits would be predicted correctly.[12] This is substantially lower than Professor Hoxby's model, which accurately predicts more than 72% of admits. Again, notwithstanding the flaws in Professor Hoxby's model (in particular her failure to include key variables as controls, see Section 2.1.3), her model still fits the data very well.

2.2.3 Professor Hoxby's misunderstanding and misapplication of Pseudo R-Squares pervade and undermine her analysis of race-neutral alternatives.

Professor Hoxby's misapplication of pseudo-R squares affect other areas of her report. Namely, Professor Hoxby argues that race-neutral admissions processes will necessarily be less effective at providing racial diversity because they will rely on proxies that are imperfect substitutes for race itself. To support her argument, Professor Hoxby presents two models that attempt to predict URM status based only on socioeconomic variables.[13] The first model uses the American Community Survey (ACS) from the U.S. Census Bureau while the second uses the NCERDC data on N.C. public high school students.

Professor Hoxby misinterprets the Pseudo R-square for these models as well. Two examples are:

(with regard to the ACS model)

---

12 Note that this also explains why the model is more accurate for rejections than admissions. It is easier to get the prediction right for the outcome that occurs more often.

13 See Section V.2, Hoxby Report, Pages 51-55.

> A model based on the entire U.S. population explains 12 percent of the variation in URM status (i.e. URM versus not). In other words, the model produces a statistic that is only 12 percent sufficient or, in other words**, able to predict URM status accurately 12 percent of the time. [emphasis added]** (Hoxby report page 53)
>
> (with regard to the NCERDC model)
>
> A regression based on all students, regardless of academic preparation, can **predict URM status with 17 percent accuracy**. That is, the regression can generate a statistic that is at most 17 percent sufficient (R-squared is 17 percent). This is not a high level of sufficiency: **predicted URM status is wrong 83 percent of the time. [emphasis added]** (Hoxby report page 54)

As I discussed in the previous two sections, Professor Hoxby's use of the Pseudo R-square is incorrect and vastly understates the accuracy of her model.

2.3 Professor Hoxby significantly understates the importance of race in her model.

Not only does Professor Hoxby's richest additive model fit the data well, but it also confirms the existence of substantial racial preferences—racial preferences that play a significant role in the admit rates of under-represented minorities. To illustrate this, Table 2.2 shows two sets of predicted admission probabilities. The first column uses Professor Hoxby's model to calculate the average predicted admission probability for African-American and Hispanic applicants with racial preferences in place. These predicted probabilities virtually match the actual admissions rates observed in the data. The second column repeats this calculation but turns the effects of race in the model off. This calculation shows how African-American and Hispanic admit probabilities would change if these applicants were instead treated as white applicants under Professor Hoxby's model. The final column is the difference between the average admit probabilities with racial

preferences and the average admit probabilities without racial preferences—what is known as "the marginal effect" of race.

Hispanic applicants have an average admit rate of over 27.9%. Using Professor Hoxby's model, the marginal effect of race for Hispanic applicants is 11.2%, implying that *over 40% of Hispanic admissions are the result of racial preferences*.

The results are even more striking for African-American applicants. African-American applicants have an average admit rate of over 24%; per Professor Hoxby's model, the admit rate of African-American applicants absent racial preferences would be less than 12%. That is, using Professor Hoxby's model, *racial preferences are responsible for more than half of African-American admissions*. Put another way, Professor Hoxby's estimates show that *racial preferences double the number of African-American admits*.

Table 2.2: Professor Hoxby's additive model shows strong racial preferences

	Average Admission Probability			
	with Racial Preferences	without Racial Preferences	Marginal Effect of Race	**Share due to Racial Preferences**
African Americans	24.3%	11.7%	12.6%	**51.8%**
Hispanics	27.9%	16.6%	11.2%	**40.3%**

Note: Predicted probabilities are generated from Professor Hoxby's additive model 9. Marginal effect is the difference between the average admission probability with racial preferences and the average admission probability without racial preferences. Share due to racial preferences is calculated as the marginal effect of race divided by the admit rate with racial preferences.

It is important to note that, as I discussed in Section 2.1.3, Professor Hoxby's models miss key components of the analysis that tend to understate the effect of racial preferences. These include important controls such as the ratings of the applicants. These also include the heterogeneous way that racial preferences operate for different applicants, such as being stronger for out-of-state applicants and applicants who do not have FGC status.

2.4 Professor Hoxby makes other important errors in her construction of the data set and in her choice of controls.

While the errors I have discussed so far are the most serious, Professor Hoxby makes other technical errors. And there are further areas where we disagree on the appropriate way to use data.

To begin, I impose a number of restrictions to the data that Professor Hoxby does not. First, I remove applicants who withdrew their applications or submitted incomplete applications. All of these applicants were rejected. These observations should clearly be removed from the analysis because no matter what their characteristics were (e.g. test scores, grades) they would be rejected. Over the six admissions cycles, 5.3% of applicants fall into this category.

Second, I focus on domestic applications as there is no claim made by the plaintiffs regarding the role of race for foreign applications. Third, I removed from the data those who were in recruiting categories with especially high admission rates. These included recruited athletes as well as those who were recruited for particular scholarships. As I noted in my initial report, applicants in these categories had admit rates of over 97%.

Here it is important to note what excluding particular applications means from a modeling perspective. Namely, it allows the admissions process to operate differently for those excluded from the analysis without specifying *how* the admission process operates differently. For example, test scores may be less important for recruited athletes and ability to pay may be more important for foreign applicants.[14] Removing these observations is implicitly

[14] Excluding these applicants is equivalent to Professor Hoxby's multiplicative model where the excluded groups are interacted with all the controls. I do not include all these interactions because there are not enough observations in each of these groups to yield meaningful results.

allowing for more flexibility in the model as well as focusing on the set of applicants where racial preferences are relevant.

Professor Hoxby's approach is instead to include as controls indicators for foreign applicants (and sometimes country of residence) and athletes.[15] This assumes that racial preferences—as well as how other control variables affect admission—operate in the same way for these groups as for domestic applicants who do not fall within one of the special recruiting categories.[16]

Fourth, while Professor Hoxby and I agree that whether an applicant applied through early action or regular action should be in the model, Professor Hoxby miscoded this variable in her analysis. Professor Hoxby created a variable that is designed to capture whether the applicant applied as part of the early action process, but an error in her analysis results in all applicants being coded as though they applied under regular action.

Fifth, in contrast to Professor Hoxby, I estimate separate models for out-of-state and in-state (North Carolina) residents. I do this because the admissions process operates very differently for these two groups. For example, deposition testimony states that legacy preferences are larger for out-of-state applicants, a result confirmed in my empirical analysis. Further, racial preferences, although large for both in-state and out-of-state applicants, are significantly larger for out-of-state applicants.[17]

---

15 Professor Hoxby does not distinguish recruited athletes from non-recruited athletes.

16 Professor Hoxby's multiplicative model allows for interactions between URM status and all of the other controls, but not interactions between the other controls and special recruiting categories and foreign applicants. Note, too, that my evidence shows differences within URM applicants: racial preferences operate differently for African-American and Hispanic applicants.

17 In principle, Professor Hoxby's multiplicative model could capture these features. But this model is restricted to only doing interactions with URM status when in fact preferences operate very differently for African-American and Hispanic applicants.

Sixth, I control for first-generation college (FGC) student and interactions with this variable and race. My analysis shows that FGC applicants receive a substantial preference in the admissions process and a much larger preference than, for example, those who receive a fee waiver for their application. And, as with racial preferences, the FGC preference is larger for out-of-state applicants than in-state applicants. *However,* African-American applicants do not receive the full FGC preference. In particular, in-state African-American applicants receive only half the preference for FGC status; and out-of-state African-American applicants who are FGC receive no preference for their FGC status.

Seventh, the number of applications UNC receives in a year varies over time. With a fixed number of seats, this implies that admission rates must also vary: in years when applications are down, admission rates will need to be higher to fill the same number of seats. Both my models and Professor Hoxby's models are effectively threshold models: if an applicant has a combination of observed and unobserved characteristics that is above some threshold, then that applicant is admitted. When the applicant pool becomes more competitive, the threshold for admission rises. Including indicator variables for each admissions cycle accounts for this effect. Professor Hoxby fails to account for the different admissions thresholds over time.

Accounting for changes in the admissions threshold is also important for ensuring that the predicted number of applicants admitted matches what is seen in the data. Ignoring this effect would result in the predicted number of admits being too low in years where applications are down and too high when applications are up. This is important not only for matching the actual admission rates, but also in how to construct counterfactual admission rates, where one may assume, for example, the absence of racial preferences.

Eighth, unlike Professor Hoxby, I do not incorporate the parental education variable into my main analysis. This variable is coded differently across years, limiting its usefulness. To illustrate this, Table 2.4 takes

Professor Hoxby's construction of the parental education variable and shows how it varies across application cycles. Clearly, there is a significant change in how this variable was coded in 2014 where the coding suggests 85% of the applicants' parents had a technical college as their highest level of parental education. Further, there is an implausible shift in the fraction assigned to the "some college" category versus the "H.S. diploma" category after 2013. I instead use whether the applicant is a first-generation college student. This variable—included in the dataset—appears to be coded consistently across years: the share of FGC applicants ranges from a high of 16.9% to a low of 15.7% across the six admissions cycles I use.

Table 2.4: Parental education varies in non-credible ways over time

Parental Education Variables, By Year

	2012	2013	2014	2015	2016	2017
A-Not Indicated	11.2%	9.6%	0.0%	0.0%	0.0%	0.0%
B-Less Than HS Graduate	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
C-HS Graduate or Equivalent	6.1%	6.5%	2.6%	2.4%	2.7%	2.9%
D-Some College	5.1%	4.8%	9.9%	8.0%	8.2%	7.9%
E-Technical School	4.3%	4.5%	84.6%	8.4%	7.9%	8.6%
F-2-Year College Degree	1.0%	0.9%	2.8%	2.2%	2.2%	3.3%
G-Bachelor's Level Degree	67.1%	68.5%	0.2%	45.6%	46.0%	44.2%
H-Some Graduate School	0.0%	0.0%	0.0%	33.4%	33.1%	33.0%
N	29,497	30,835	31,331	31,956	35,875	40,918

Note: Year refers to year graduated high school.

Finally, I make use of all the admissions cycles that contain UNC's ratings, those for the graduating classes of 2012 to 2017. Professor Hoxby's analysis does not use the information from the 2012 and 2013 admissions cycles.

3. *Positive Analysis*

3.1 *Factors I incorporate from Professor Hoxby's analysis*

As discussed in Section 2, Professor Hoxby and I differ in identifying the relevant dataset as well as what controls to include. Considering my

original approach as well as Professor Hoxby's analysis, the only changes I believe are appropriate to make to my analysis are:

1) Include controls for the applicant's rank among those who have applied to UNC from the applicant's high school as well as a measure for whether this rank is missing.

2) Include a control for whether the applicant is a child of a faculty member.

3) Add a control for those who do not meet the minimum requirements of admission to UNC.

I also continue to use my controls for test scores and grades rather than Professor Hoxby's. However, this decision is more of a judgment call. Hence, in Section 3.4.3, I show that my analysis yields similar results using her controls for these variables.

3.2 The main findings of my model continue to hold.

The modeling changes described above have little effect on my analysis. In Appendices A and B, I reproduce all of the tables from my opening report that change as a result of the additional controls.[18] As those tables illustrate, the conclusions of my original report hold. In particular:

- Non-URM applicants have significantly stronger qualifications than URM applicants to UNC. (This is true for both in-state and out-of-state applicants.) In particular, non-URM applicants have significantly higher test scores and high school grades and receive higher program, performance, extracurricular, and essay ratings from UNC admissions officers than URM applicants.

- Significant racial preferences are given to in-state URM applicants over their non-URM counterparts. Racial preferences are even larger for out-of-state URMs.

- Although UNC also gives preferences to FGC students, these are much smaller than the racial preferences for URMs. Further, UNC gives

[18] Those in the body of my original report are in Appendix A; those in the appendix of my original report are in Appendix B.

first-generation URM applicants a much smaller preference for their FGC status than it gives first-generation non-URM applicants. Relative to the non-URM applicant, then, the overall preference the URM applicant receives would be larger if the two applicants were non-FGC than if they were FGC.

- To put the magnitude of racial preferences into perspective, consider a male, non-FGC Asian-American applicant whose observed characteristics would imply a 25% chance of admission:
    - If he was an in-state applicant, his probability of admission would increase to over 67% (i.e., more than double) had he been treated like an in-state Hispanic applicant, and to over 90% (more than triple) had he been treated like an in-state African American applicant across the full sample period.
    - If he was an out-of-state applicant, his probability of admission would increase to over 86% had he been treated like an out-of-state Hispanic applicant and over 99% had he been treated like an out-of-state African-American applicant.
    - Simply changing this hypothetical Asian-American applicant's race to either Hispanic or African-American thus would transform him from an unlikely admit to an almost certain admit.
- Holding fixed the number of admission slots and removing racial preferences would significantly change Asian-American and white representation at UNC. Fixing the number in-state admits at the levels observed in the data and removing racial preferences would result in 1,171 additional non-URM admits over the six-year period, a 5.3% increase. Fixing the number of out-of-state admits at the levels observed in the data and removing racial preferences would result in 2,486 more non-URM admits from that pool, a 25.8% increase.

3.3 My preferred model does an excellent job fitting the data.

I now examine how well the observables alone do at predicting admissions decisions using my preferred model (Model 4 in Appendix B, Table A.4.1.R and A.4.2.R.). Specifically, I use the coefficients of my preferred in-state and out-of-state models and the observed characteristics of each of the applicants to form an *admissions index* similar to what I did with

Professor Hoxby's models in Section 2.2.2. For each admissions cycle and separately for out-of-state and in-state applicants, I rank the applicants according to their admissions index. Using this ranking, I then assign those with the highest rankings as those who would be predicted to be admitted based on the observables alone such that the number of admitted students matches what is observed for that admissions cycle for both in-state and out-of-state admits.

3.3.1 UNC's in-state admissions are guided by an implicit formula that accurately predicts over 90% of admission decisions.

Table 3.1 shows how well my preferred model predicts who is admitted from the in-state applicant pool according to the part of the formula associated with observed characteristics. *Over 91%* of those who are actually admitted in the in-state applicant pool would be predicted to be admitted based on their observables alone. And over 92% of those who are actually rejected in the in-state applicant pool are predicted to be rejected based on their observables. Since admission rates are close to 50% in the in-state pool, this would roughly be the rate of correct predictions if students were randomly assigned to fill the admissions slots. As shown in the third column, the overall accuracy of the model—the total number of correct predictions divided by the number of applicants—is also over 91%.

I should note that these types of numbers are virtually never reported in economics papers, primarily because most models would not provide nearly as accurate predictions. The reason for this is two-fold. First, the discovery process has yielded highly detailed data on UNC applicants. Second, UNC's admissions process uses the observed characteristics in a way that leaves little room for unobserved factors (in other words, the unobserved factors have little effect on admissions decisions).

Table 3.1: Accuracy of my preferred model for in-state admissions

	Accuracy for Admits	Accuracy for Rejects	Overall Accuracy
Preferred Model	91.8%	92.5%	92.1%
Random Assignment	48.1%	52.2%	50.2%

Note: My preferred model refers to column 4 of Appendix B, Table A.4.1.R for the in-state analysis. Accuracy is defined as the number of correct predictions divided by the appropriate sample size. For example, accuracy for admits is defined as the number of admits who were predicted to be admitted divided by the number of admits; overall accuracy is the number of admits who were predicted to be admitted plus the number of rejects predicted to be rejected divided by the number of applicants.

A key point of this analysis, then, is that UNC's undergraduate admissions process is very formulaic: the various characteristics of applicants can be assigned numerical scores and added up; those who are above some threshold are admitted, and the rest are rejected. Of course, this is not literally true: UNC admissions officers undoubtedly look at information that was not part of my analysis (e.g., letters of recommendation) and debate among themselves the merits of particular borderline candidates. But the formulaic elements I have modeled dominate the actual outcomes.

I present a way of visualizing this, for the in-state applicants, in Figures 1 and 2.[19] My model generates a prediction (from the "inputs" of the admissions process, like test scores and high school grades) of the probability of each student's admission. Figure 1 plots where those predictions fall along a continuum from "0" (certain to be rejected) to "1" (certain to be admitted). The height of the curve in Figure 1 corresponds to how many of the predictions fall at particular points along the 0 to 1 range. As the reader can see, the predictions are heavily clustered toward the "0" end of the range, and the "1" end of the range, with very few predictions in the middle. This means that the model has an easy time sorting applicants into those very likely to be

19 In Figure 1 and 2, I plot the distribution of the predicted probabilities for my preferred model using *kernel density estimation* (KDE). The output of KDE is essentially the equivalent of a histogram except that the underlying data is continuous.

rejected and very likely to be admitted. Because the curve is so "U-shaped", with such a shallow middle section, means there were very few "borderline" cases resolved by factors not in my model.

Figure 2 shows how well the predictions of the model work out when compared with actual admissions decisions. The red line shows how the "predicted probabilities of admission" were distributed for admitted students. As the reader can see, the vast majority of students actually admitted were, in the model, estimated to have very high probabilities of admission (almost always above 90%). Conversely, the blue line shows the "predicted probabilities of admission" for the students actually rejected. Nearly all of the students actually rejected were, in the model, estimated to have very low probabilities of admission (almost always below 10%). In other words, the predictions of the model line up with real outcomes to an extraordinary degree.

I should note that even in a system that is not formulaic, a researcher with some objective data about the applicants can sometimes predict with confidence that certain applicants will be *rejected*. That is because all selective schools have some threshold of minimum qualifications below which they are reluctant to venture. Students with qualifications below that threshold thus have a low chance of admission. What is striking about these results is the extraordinary accuracy with which we can predict who is *accepted*. This shows a heavily formulaic approach in which the key factors driving admission are combined in a largely mechanical way.

Figure 1: The distribution of predicted admit probabilities for in-state applicants

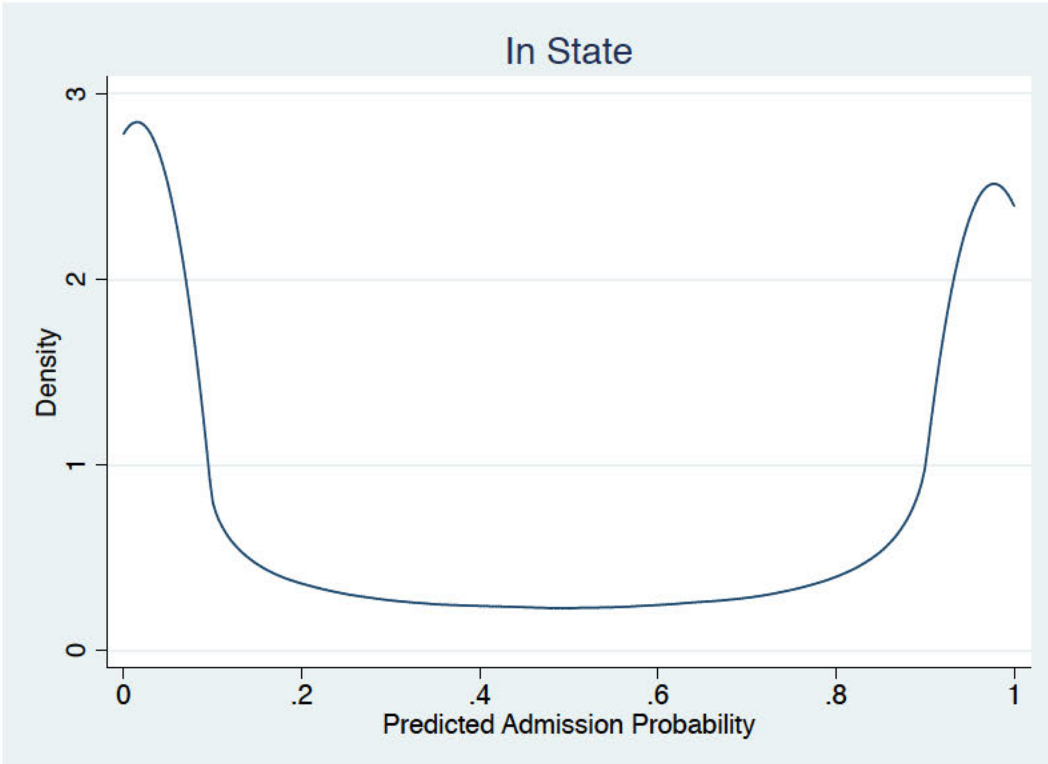
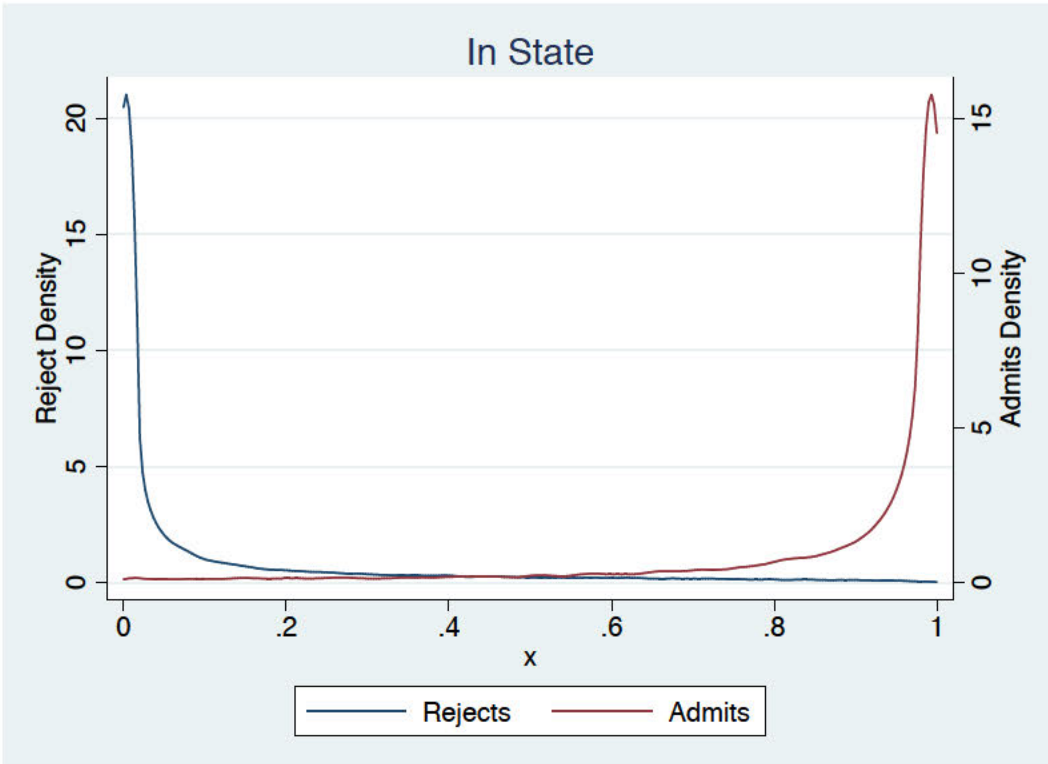


Figure 2: The distribution of predicted probabilities conditional on being rejected or admitted for in-state applicants



3.3.2 My model predicts UNC's out-of-state admissions very well.

Table 3.2 shows the same results as Table 3,1 but now for out-of-state applicants. Over 75% of those who are actually admitted in the out-of-state applicant pool would be predicted to be admitted based on their observables alone. While this number is lower than the corresponding in-state rate, this is not surprising because the average admit rate for out-of-state applicants is substantially lower that in-state rate: if applicants were randomly assigned to fill the out-of-state admissions slots in each year, only 13.9% of admits would be correctly assigned, implying that my model produces an over five-fold increase in the probability of a correct prediction over random assignment. As rejections are more common in the out-of-state sample, the model correctly predicts over 96% of rejections, resulting in an overall accuracy of the model of over 93%.

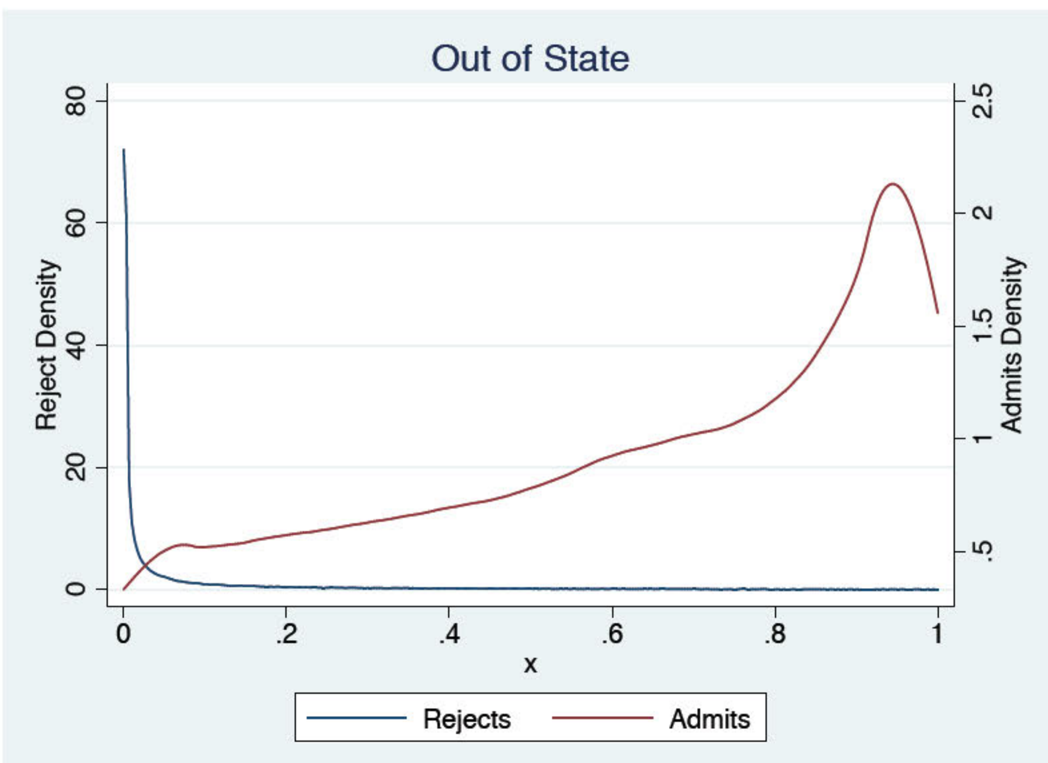
Table 3.2: Accuracy of my preferred model for out-of-state admissions

	Accuracy for Admits	Accuracy for Rejects	Overall Accuracy
Preferred Model	75.4%	96.1%	93.3%
Random Assignment	13.9%	86.5%	76.7%

Note: My preferred model refers to column 4 of Appendix Table A.4.2.R. for the out-of-state analysis. Accuracy is defined as the number of correct predictions divided by the appropriate sample size. For example, accuracy for admits is defined as the number of admits who were predicted to be admitted divided by the number of admits; overall accuracy is the number of admits who were predicted to be admitted plus the number of rejects predicted to be rejected divided by the number of applicants.

Figure 3 repeats the analysis in Figure 2 except now for the out-of-state sample. Because the out-of-state admit rate is only 13.9% over the six-year period, predicting admissions is more difficult. Nonetheless, Figure 3 shows that the probability mass for those who are rejected is around zero while it is around one for those who are admitted.

Figure 3: The distribution of predicted probabilities conditional on being rejected or admitted for out-of-state applicants



3.4 Racial preferences at UNC are very large.

3.4.1 If African-American and Hispanic applicants were treated as white or Asian-American applicants, their admit rates would fall substantially.

As I described in my previous report and at the beginning of Section 3, my estimates of racial preferences are substantial. Here, I present another way of seeing this by highlighting how much of the African-American and Hispanic admit rate is the result of racial preferences. This analysis mirrors what I reported for Professor Hoxby's results, but because I model out-of-state and in-state admissions separately, the results are by residency status.

Each panel of Table 3.3 shows two sets of predicted admission probabilities, with the first (second) panel showing results for in-state (out-of-state) applicants. The first column uses my preferred model to calculate the average predicted admission probability for African-American and Hispanic applicants with racial preferences in place.[20] The second column repeats this calculation but turns the effects of race in the model off. This calculation shows how African-American and Hispanic admit probabilities would change if these applicants were instead treated as white applicants under my preferred model. The final column is the difference between the average admit probabilities with racial preferences and the average admit probabilities without racial preferences, the marginal effect of race.

In-state Hispanic applicants are admitted at a rate of 41% and would experience a 9.7% drop in their admission rate if they were treated as white applicants. For African-American applicants, the estimated preferences are substantially larger. In-state African-American applicants are admitted at a

[20] My preferred models show that some characteristics result in "perfect predictions" where, for example, a very low performance rating may guarantee rejection regardless of race. I do not include perfect predictions in my calculations as racial preferences have no bearing for these applicants. Hence these predictions are only for those who have characteristics that do not automatically result in admission or rejection regardless of race.

rate of 30.5% with racial preferences. Removing racial preferences results in a drop in admissions rate of 12.7%. In other words, racial preferences account for nearly 42% of African-American in-state admissions.

The marginal effects are even larger for out-of-state applicants notwithstanding the fact that the base admission rates are much lower. For out-of-state Hispanic applicants, the numbers are stark: with UNC's racial preferences, their admit rates are 20.3%; without racial preferences, their admit rates would fall to 6.0%. Racial preferences thus account for 70% of out-of-state Hispanic admissions. The results are even more striking for out-of-state African-American applicants. With UNC's racial preferences, their admit rates are 17%; *without racial preferences, their admit rates would fall to 1.5%. Racial preferences thus account for 91% of out-of-state African-American admissions.*

Table 3.3: My preferred models shows strong racial preferences in both the in-state and out-of-state admissions processes.

	Average Admission Probability			
	with Racial Preferences	without Racial Preferences	Marginal Effect of Race	**Share due to Racial Preferences**
In-state				
African American	30.5%	17.8%	12.7%	**41.7%**
Hispanic	41.0%	31.2%	9.7%	**23.8%**
Out-of-state				
African American	17.1%	1.5%	15.6%	**91.1%**
Hispanic	20.3%	6.0%	14.2%	**70.2%**

Note: My preferred models refer to column 4 of Appendix Table A.4.1.R for the in-state analysis and column 4 of Appendix B Table A.4.2.R for the out-of-state analysis. Marginal effect is the difference between the average admission probability with racial preferences and the average admission probability without racial preferences. Share due to racial preferences is calculated as the marginal effect of race divided by the admit rate with racial preferences.

In my previous report, I showed that, as compared with non-URM applicants, the preferences UNC gives to URM applicants are much stronger for non-FGC applicants. This is because FGC applicants who are URMs do

not receive the full preference for their FGC status. I now show how the marginal effects of racial preferences differ between those with and those without FGC status.

Results are presented in Table 3.4. These results show that UNC uses racial preferences to greater effect for non-FGC URM applicants than FGC URM applicants. I focus my discussion on African-American applicants where the differences are most pronounced. For in-state African-American applicants, the average marginal effect of racial preferences for non-FGC applicants is 14.9 percentage points. This is over fifty percent higher than the average marginal effect of race preferences for African-American FGC applicants that was 9.3 percentage points. The share of admissions that are the result of racial preferences is 44.8% for African-American applicants who were not FGC compared to 35.4% for those who were FGC.

The second panel shows the results for out-of-state applicants where the gap in the marginal effect of racial preferences between FGC and non-FGC URM applicants is even larger. The marginal effect of racial preferences for African-American applicants who were not FGC was 17.5 percentage points compared to 10.5 percentage points for those who were FGC. The share of admissions that are the result of racial preferences is 91.7% for African-American applicants who were not FGC compared to 88.6% for those who were FGC.

Table 3.4: Racial preferences are smaller for first generation college students

	Average Admission Probability with Racial Preferences	without Racial Preferences	Marginal Effect of Race	Share due to Racial Preferences
In-state				
African American non-FGC	33.4%	18.4%	14.9%	44.8%
African American FGC	26.1%	16.9%	9.3%	35.4%
Hispanic non-FGC	45.8%	34.7%	11.0%	24.1%
Hispanic FGC	35.5%	27.2%	8.3%	23.3%
Out-of-state				
African American non-FGC	19.1%	1.6%	17.5%	91.7%
African American FGC	11.9%	1.4%	10.5%	88.6%
Hispanic non-FGC	22.1%	6.2%	15.9%	72.0%
Hispanic FGC	13.7%	5.5%	8.2%	59.7%

Note: My preferred models refer to column 4 of Appendix B, Table A.4.1.R for the in-state analysis and column 4 of Appendix Table A.4.2.R. for the out-of-state analysis. Marginal effect is the difference between the average admission probability with racial preferences and the average admission probability without racial preferences. Share due to racial preferences is calculated as the marginal effect of race divided by the admit rate with racial preferences.

3.4.2 Factors not accounted for by the implicit formula are significantly less important than racial preferences for African-American and Hispanic applicants.

As the previous section showed, racial preferences have a substantial effect on the admission rates of under-represented minorities. In this section, I show that racial preferences are substantially more important than unobserved factors and thus that race has an outsize impact on admissions decisions. To do this, I use three pieces of information: the distribution of the unobserved characteristic, the predicted probability of admission, and the actual admission decision. With this information, I show for admitted African-American and Hispanic applicants:[21]

- How often the expected value of the unobserved characteristic is larger than the preferences for a particular racial/ethnic group; and
- The probability of the unobserved characteristic being larger than the preferences for a particular racial/ethnic group.

[21] The derivations of these formulas are in Appendix D.

The first panel of Table 3.5 shows the results for in-state applicants. There are two important points to be made here, both of which demonstrate that racial preferences matter much more than unobserved characteristics. First, for nearly every URM admit, the magnitude of UNC's racial preferences are larger than the expected values of their unobserved characteristics. This demonstrates that racial preferences are substantially more important to URM admissions than unobserved factors (both in-state and out-of-state). This can be seen in the first column of Table 3.5, which shows that fewer than 2% of in-state Hispanic admits have expected values of their unobserved characteristics that are larger than the racial preferences they receive. (Put another way, racial preferences matter more than the expected value of the unobserved factors for more than 98% of in-state Hispanic admits.). Similarly, fewer than 1% of in-state African-American admits have expected values of their unobserved characteristic that are larger than the racial preferences they receive. (Again, this means that racial preferences are larger than the expected value of the unobserved factors for more than 99% of in-state African-American admits.)

The second panel of Table 3.5 demonstrates the same for out-of-state applicants. Fewer than 3% of out-of-state Hispanic admits have expected values of the unobserved characteristics that are larger than the racial preference they receive. Similarly, fewer than 0.2% of African-American admits have expected values of the unobserved characteristic that are larger than the racial preference they receive. Racial preferences thus are substantially more important to URM out-of-state admissions too.

Second, the average probability that an admit's unobserved characteristics carry more weight than racial preferences is very low. This is illustrated in the second column of Table 3.5. For in-state applicants, the average probability that a Hispanic admit's unobserved characteristics carry more weight than racial preferences is 17.9%; for African-American admits, that probability is only 5.6%. For out-of-state applicants, these figures are

even lower: the average probability that a Hispanic admit's unobserved characteristics carry more weight than racial preferences is 13.8%; for African-American admits, that probability is only 1.2%.

Table 3.5: Racial Preferences are Substantially More Important to URM Admissions than Unobserved Factors

	Fraction with Expected Unobserved Factor Greater than Racial Preference	Average Probability that Unobserved Factor is Larger than Racial Preferences
In-State		
African-American admits	0.66%	5.64%
Hispanic Admits	1.71%	17.94%
Out-of-State		
African-American admits	0.12%	1.17%
Hispanic Admits	2.56%	13.76%

Note: See Appendix D for formulas. Calculated using Model 4 in Appendix B, Table A.4.1.R. for in-state applicants and Appendix B, Table A.4.2.R. for out-of-state applicants..

3.4.3 My estimates of large racial preferences are robust to alternative ways of coding grades and test scores.

As explained above, Professor Hoxby's modeling suffers from a number of flaws that undermine her analysis. In other instances, she made different modeling choices than I did. Some of these choices were judgment calls where legitimate disagreements could exist across researchers. In this section, I show that my model is robust to those choices—that is, making the choices Professor Hoxby made would not change the results yielded by my model. In particular, Professor Hoxby coded test score and grade information in a different way than I did. To show that my results are robust to these coding choices, I replace my way of coding test score and grade information with Professor Hoxby's approach.

I use Professor Hoxby's measures in two ways. First, I add the controls as she constructed them subject to one minor adjustment.[22] Second, I add interactions between missing test score and grade variables with race. Including interactions between missing test scores and race is appropriate because, absent these measures, the same average test score is imputed for all applicants who are missing these data. The interactions allow for different average test scores to be imputed for different racial groups, taking into account the large differences in test scores across races in the UNC applicant pool.[23] The estimated results are presented in Appendix C, Tables C.1 and C.2 for the in-state and out-of-state data respectively.

To see how the different approaches to these variables affect my estimates of the significance of race in the admissions process, Table 3.6 reports marginal effects of race for both the in-state and out-of-state models. The estimated marginal effects of my preferred model lie in between the estimated marginal effects using the two ways of incorporating Professor Hoxby's measures of grades and test scores—one without interactions between missing tests and race and one with those interactions Namely, without missing race interactions, the estimated marginal effects are slightly lower than my preferred model (though still substantial); with missing race interactions, the effects are larger than my preferred model.

---

22 The one adjustment is an interaction between year and missing class rank. I included this interaction in my analysis as well. I do this because the rate at which class rank is missing varies substantially by year. In the first year of the data, 29% of domestic applicants are missing class rank while 46% are missing class rank in the last year.

23 As discussed in my previous report, interactions between missing variables and race are also included in my model. Arcidiacono Report 40.

Table 3.6: Strong racial preferences are robust to alternative constructions of test scores and grades

	Average Admission Probability with Racial	Marginal Effect of Race		
		Preferred Model	Robustness (1)	Robustness (2)
In-state				
African American	30.5%	12.7%	11.9%	14.2%
Hispanic	41.0%	9.7%	8.9%	11.5%
Out-of-state				
African American	17.1%	15.6%	14.9%	16.4%
Hispanic	20.3%	14.2%	13.7%	15.7%

Note: My preferred models refer to column 1 of Appendix Table C.1 for the in-state analysis and column 1 of Appendix Table C.2 for the out-of-state analysis. Robustness (1) replaces my measures of test scores, grades, and class rank with Professor Hoxby's. Robustness (2) additionally adds interactions between missing these variables and race. Estimates of Robustness (1) and (2) are given in Appendix Table C.1 and Table C.2. Marginal effect is the difference between the average admission probability with racial preferences and the average admission probability without racial preferences.

Dated: April 6, 2018

<u>/s/ Peter S. Arcidiacono</u>
Peter S. Arcidiacono

Appendix A

(Tables included in Appendix A are labeled to correspond with the tables in the body of my opening report)

Table 2.3.R: Application Summary Statistics by Race: In-State Applicants

	White			African American			Hispanic			Asian American			Total		
	Reject	Admit	Total	Reject	Admit	Total	Reject	Admit	Total	Reject	Admit	Total	Reject	Admit	Total
Admitted	0.00	100.00	50.86	0.00	100.00	30.53	0.00	100.00	40.96	0.00	100.00	53.56	0.00	100.00	47.92
Female	56.93	60.67	58.83	65.91	70.09	67.19	62.25	61.22	61.83	55.83	56.97	56.44	58.87	61.10	59.94
First-generation college	18.25	13.21	15.69	41.68	33.57	39.20	51.06	40.48	46.73	30.03	20.04	24.68	26.06	17.27	21.85
Legacy	17.42	21.84	19.67	6.35	9.27	7.24	4.53	4.90	4.68	4.62	5.77	5.24	13.09	17.76	15.33
Waiver	6.92	5.06	5.97	45.86	37.99	43.46	36.01	28.91	33.10	16.26	12.19	14.08	17.34	10.27	13.95
Faculty Child	1.44	1.91	1.68	1.07	0.93	1.03	1.32	1.16	1.25	2.08	2.98	2.56	1.42	1.91	1.65
Min Req Not Met	0.13	0.01	0.07	0.13	0.00	0.09	0.28	0.00	0.17	0.04	0.00	0.02	0.14	0.00	0.07
Missing SAT or ACT	0.29	0.00	0.14	0.59	0.00	0.41	0.71	0.07	0.45	0.36	0.00	0.17	0.38	0.00	0.20
Missing class percentile	14.53	16.20	15.38	7.42	13.14	9.17	9.91	13.81	11.51	9.09	18.18	13.96	12.45	16.40	14.34
Missing GPA	0.03	0.04	0.04	0.02	0.04	0.03	0.03	0.04	0.03	0.03	0.06	0.04	0.03	0.04	0.04
Missing Rank w/in HS applicants	14.75	16.48	15.63	7.72	13.69	9.54	10.24	13.88	11.73	9.27	18.43	14.18	12.68	16.70	14.61
SAT math (z-score)	-0.69	0.06	-0.31	-1.57	-0.73	-1.31	-1.17	-0.37	-0.84	-0.46	0.47	0.04	-0.87	0.02	-0.44
	(0.75)	(0.66)	(0.80)	(0.82)	0.69	(0.87)	(0.83)	(0.73)	(0.88)	(0.92)	(0.73)	(0.95)	(0.87)	(0.74)	(0.93)
SAT verbal (z-score)	-0.55	0.25	-0.14	-1.42	-0.51	-1.14	-1.03	-0.16	-0.68	-0.74	0.27	-0.20	-0.77	0.17	-0.32
	(0.83)	(0.73)	(0.88)	(0.88)	0.81	(0.96)	(0.93)	(0.83)	(0.99)	(0.96)	(0.84)	(1.03)	(0.93)	(0.80)	(0.98)
High school class percentile (0-100)	79.10	93.58	86.34	74.20	91.56	79.26	74.87	91.81	81.57	74.86	92.78	83.94	77.28	93.14	84.64
	(14.21)	(6.17)	(13.13)	(17.51)	(7.80)	(17.24)	(16.54)	(7.46)	(16.00)	(16.26)	(6.96)	(15.33)	(15.48)	(6.63)	(14.54)
GPA (z-score)	-0.09	0.90	0.41	-0.67	0.59	-0.28	-0.39	0.74	0.07	-0.14	1.03	0.49	-0.23	0.87	0.30
	(0.78)	(0.60)	(0.85)	(0.90)	(0.57)	(1.00)	(0.82)	(0.58)	(0.92)	(0.82)	(0.59)	(0.92)	(0.84)	(0.61)	(0.92)
Rank w/in HS Applicants	34.73	72.33	53.66	26.41	62.18	36.83	28.60	64.95	43.13	32.88	74.74	54.19	32.36	71.15	50.49
	(19.65)	(17.37)	(26.40)	(19.01)	(19.54)	(25.13)	(19.14)	(19.28)	(26.18)	(19.73)	(16.76)	(27.79)	(19.76)	(18.06)	(27.11)
Program Rating (1-10)	5.35	7.52	6.45	4.94	7.30	5.66	5.34	7.42	6.19	6.35	8.56	7.53	5.37	7.62	6.45
	(2.32)	(2.00)	(2.42)	(2.62)	(2.31)	(2.75)	(2.56)	(2.18)	(2.62)	(2.58)	(1.76)	(2.44)	(2.45)	(2.05)	(2.53)
Performance Rating (1-10)	5.62	8.37	7.02	4.55	7.40	5.42	4.93	7.77	6.09	4.96	8.03	6.60	5.29	8.20	6.69
	(1.89)	(1.41)	(2.16)	(1.82)	(1.60)	(2.19)	(1.85)	(1.51)	(2.22)	(1.89)	(1.52)	(2.29)	(1.92)	(1.48)	(2.26)
Extracurricular Rating (1-10)	5.42	6.07	5.75	4.90	5.69	5.15	5.00	5.71	5.29	5.09	6.02	5.59	5.26	6.01	5.62
	(1.20)	(1.03)	(1.16)	(1.39)	(1.10)	(1.36)	(1.37)	(1.17)	(1.34)	(1.35)	(1.14)	(1.33)	(1.28)	(1.07)	(1.24)
Essay Rating > 5	0.05	0.15	0.10	0.03	0.13	0.06	0.04	0.14	0.08	0.05	0.19	0.13	0.05	0.16	0.10
	(0.22)	(0.36)	(0.31)	(0.18)	(0.33)	(0.24)	(0.20)	(0.35)	(0.27)	(0.23)	(0.39)	(0.33)	(0.21)	(0.36)	(0.30)
Personal Quality Rating > 5	0.11	0.24	0.18	0.14	0.32	0.20	0.17	0.32	0.23	0.11	0.27	0.20	0.12	0.26	0.19
	(0.31)	(0.43)	(0.38)	(0.35)	(0.47)	(0.40)	(0.38)	(0.47)	(0.42)	(0.32)	(0.45)	(0.40)	(0.33)	(0.44)	(0.39)
N	18,229	18,865	37,094	5,401	2,374	7,775	2,119	1,470	3,589	2,794	3,223	6,017	29,803	27,422	57,225

Table 2.4.R: Application Summary Statistics by Race: Out-of-State Applicants

	White			African American			Hispanic			Asian American			Total		
	Reject	Admit	Total	Reject	Admit	Total	Reject	Admit	Total	Reject	Admit	Total	Reject	Admit	Total
Admitted	0.00	100.00	10.91	0.00	100.00	16.74	0.00	100.00	20.18	0.00	100.00	16.60	0.00	100.00	13.52
Female	61.18	55.26	60.53	66.00	66.79	66.13	59.26	60.35	59.48	55.73	54.74	55.56	60.53	57.64	60.14
First-generation college	8.97	7.22	8.78	29.75	19.00	27.95	23.97	14.94	22.14	13.37	8.90	12.63	12.66	9.85	12.28
Legacy	2.63	17.82	4.29	1.78	3.80	2.12	1.25	4.34	1.87	0.80	3.00	1.16	2.11	10.99	3.31
Waiver	3.70	2.63	3.58	36.39	25.67	34.60	18.54	12.63	17.35	9.92	6.63	9.37	8.80	7.42	8.61
Minimum Requirements Not Met	0.09	0.03	0.08	0.36	0.00	0.30	0.10	0.00	0.08	0.07	0.00	0.06	0.11	0.01	0.10
Missing SAT or ACT	0.20	0.01	0.18	0.69	0.00	0.57	0.42	0.11	0.35	0.15	0.00	0.12	0.25	0.03	0.22
Missing class percentile	39.68	39.55	39.66	33.06	32.52	32.97	39.49	41.35	39.86	44.33	42.55	44.04	40.16	39.75	40.10
Missing GPA	0.29	0.32	0.30	0.28	0.29	0.28	0.28	0.31	0.29	0.31	0.33	0.31	0.30	0.32	0.30
Missing Rank w/in HS Applicants	41.75	42.74	41.86	35.63	35.39	35.59	41.99	46.24	42.85	47.28	46.78	47.19	42.45	43.34	42.57
SAT math (z-score)	-0.01	0.80	0.08	-1.16	-0.08	-0.98	-0.43	0.40	-0.27	0.48	1.20	0.60	-0.06	0.73	0.04
	(0.76)	(0.54)	(0.78)	(0.93)	0.7	(0.99)	(0.86)	(0.61)	(0.88)	(0.77)	(0.41)	(0.77)	(0.89)	(0.67)	(0.90)
SAT verbal (z-score)	0.15	1.02	0.24	-0.91	0.24	-0.72	-0.25	0.64	-0.07	0.22	1.17	0.38	0.04	0.91	0.16
	(0.79)	(0.57)	(0.81)	(0.99)	0.71	(1.04)	(0.88)	(0.64)	(0.91)	(0.85)	(0.57)	(0.88)	(0.89)	(0.66)	(0.91)
High school class percentile (0-100)	87.49	96.75	88.44	77.18	94.01	79.88	83.34	95.34	85.57	86.78	97.14	88.53	85.83	96.22	87.18
	(12.74)	(4.60)	(12.47)	(18.25)	(6.36)	(18.01)	(15.47)	(5.64)	(14.92)	(13.49)	(3.92)	(13.00)	(14.22)	(5.06)	(13.83)
GPA (z-score)	-0.16	0.29	-0.11	-0.70	0.07	-0.57	-0.21	0.37	-0.09	-0.16	0.34	-0.07	-0.21	0.28	-0.15
	(0.74)	(0.62)	(0.74)	(0.99)	(0.64)	(0.99)	(0.89)	(0.73)	(0.89)	(0.75)	(0.59)	(0.75)	(0.79)	(0.64)	(0.79)
Rank w/in HS Applicants	51.07	77.68	53.93	32.96	61.03	37.68	42.46	70.04	47.69	53.62	82.44	58.44	49.00	75.23	52.50
	(27.65)	(20.12)	(28.18)	(25.59)	(24.55)	(27.49)	(27.04)	(23.19)	(28.48)	(26.66)	(16.68)	(27.46)	(27.90)	(21.71)	(28.58)
Program Rating (1-10)	6.16	8.40	6.40	5.08	7.65	5.51	6.26	8.42	6.70	7.26	9.10	7.57	6.25	8.45	6.55
	(2.56)	(2.00)	(2.60)	(2.67)	(2.31)	(2.78)	(2.80)	(1.98)	(2.79)	(2.50)	(1.58)	(2.47)	(2.64)	(2.00)	(2.67)
Performance Rating (1-10)	7.32	9.08	7.51	5.29	8.01	5.75	6.35	8.58	6.80	6.85	9.06	7.22	6.97	8.88	7.23
	(2.02)	(1.19)	(2.02)	(2.08)	(1.54)	(2.24)	(2.07)	(1.30)	(2.14)	(2.06)	(1.07)	(2.10)	(2.12)	(1.28)	(2.13)
Extracurricular Rating (1-10)	5.87	6.83	5.98	5.24	6.29	5.41	5.57	6.50	5.76	5.81	7.02	6.01	5.79	6.77	5.92
	(1.10)	(0.95)	(1.13)	(1.35)	(1.03)	(1.36)	(1.23)	(0.94)	(1.23)	(1.18)	(1.02)	(1.24)	(1.17)	(1.00)	(1.19)
Essay Rating > 5	0.11	0.44	0.15	0.08	0.30	0.12	0.10	0.35	0.15	0.14	0.50	0.20	0.11	0.43	0.16
	(0.31)	(0.50)	(0.35)	(0.27)	(0.46)	(0.32)	(0.30)	(0.48)	(0.36)	(0.35)	(0.50)	(0.40)	(0.32)	(0.49)	(0.36)
Personal Quality Rating > 5	0.16	0.53	0.20	0.19	0.51	0.24	0.20	0.50	0.27	0.17	0.56	0.24	0.17	0.54	0.22
	(0.37)	(0.50)	(0.40)	(0.39)	(0.50)	(0.43)	(0.40)	(0.50)	(0.44)	(0.38)	(0.50)	(0.43)	(0.38)	(0.50)	(0.42)
N	56,790	6,954	63,744	7,980	1,605	9,585	7,202	1,821	9,023	13,554	2,698	16,252	91,351	14,281	105,632

Table 4.1.R: The magnitude of Racial/Ethnic Preferences for In-state Applicants

	Original admit probability	Admission probability if treated as:					
		African American			Hispanic		
		Model 4	Model 6	Model 7	Model 4	Model 6	Model 7
Asian American							
Male, not FGC	25%	90.85%	93.82%	96.94%	67.85%	74.02%	79.10%
	10%	76.80%	83.50%	91.34%	41.30%	48.72%	55.78%
Female, not FGC	25%	88.83%	92.05%	95.75%	69.60%	77.70%	80.42%
	10%	72.61%	79.42%	88.24%	43.29%	53.73%	57.79%
Male, FGC	25%	80.49%	83.66%	89.53%	62.31%	64.83%	73.12%
	10%	57.89%	63.06%	74.03%	35.53%	38.06%	47.55%
Female, FGC	25%	76.76%	79.61%	85.88%	64.20%	69.26%	74.70%
	10%	52.40%	56.55%	66.97%	37.42%	42.89%	49.60%
White							
Male noFGC	25%	92.01%	94.72%	97.42%	70.98%	77.11%	81.86%
	10%	79.33%	85.68%	92.63%	44.92%	52.90%	60.06%
Female noFGC	25%	87.80%	90.54%	95.29%	67.45%	74.24%	78.68%
	10%	70.59%	76.14%	87.08%	40.86%	48.99%	55.15%
Male FGC	25%	80.48%	85.81%	89.99%	62.30%	68.53%	74.10%
	10%	57.88%	66.85%	74.99%	35.52%	42.05%	48.81%
Female FGC	25%	72.05%	76.34%	82.82%	58.33%	65.06%	70.06%
	10%	46.21%	51.82%	61.64%	31.81%	38.30%	43.82%

Table 4.2.R: The magnitude of Racial/Ethnic Preferences for Out-of-state Applicants

	Original admit probability	Admission probability if treated as:			
		African American		Hispanic	
		Model 4	Model 6	Model 4	Model 6
Asian American					
Male, not FGC	25%	99.32%	99.35%	86.10%	86.76%
	10%	97.99%	98.09%	67.37%	68.59%
Female, not FGC	25%	99.30%	99.38%	88.84%	89.37%
	10%	97.94%	98.17%	72.63%	73.69%
Male, FGC	25%	98.52%	98.74%	80.09%	81.79%
	10%	95.68%	96.31%	57.27%	59.96%
Female, FGC	25%	98.48%	98.80%	83.79%	85.22%
	10%	95.57%	96.48%	63.27%	65.77%
White					
Male noFGC	25%	99.37%	99.44%	87.00%	88.40%
	10%	98.14%	98.35%	69.05%	71.75%
Female noFGC	25%	99.42%	99.50%	90.54%	91.25%
	10%	98.28%	98.52%	76.13%	77.66%
Male FGC	25%	97.64%	98.15%	71.41%	75.23%
	10%	93.23%	94.64%	45.43%	50.31%
Female FGC	25%	97.81%	98.34%	78.12%	80.62%
	10%	93.72%	95.18%	54.34%	58.09%

Table 4.3.R: Predicted Admissions Probabilities for Asian Americans and Whites if Treated as African American or Hispanic

		Treated as African American		Treated as Hispanic		
	Original admit probability	Adjusted admit probability	Change in admission probability	Adjusted admit probability	Change in admission probability	Observations
Non-URM						
In-State						
No high school fixed effects	0.512	0.682	0.170	0.617	0.104	43,111
High school fixed effects sample	0.526	0.698	0.172	0.633	0.107	40,821
High school fixed effects	0.526	0.695	0.169	0.635	0.109	40,821
HS and Census tract fixed effects	0.535	0.722	0.187	0.649	0.114	30,271
Out-of-State						
No high school fixed effects	0.121	0.574	0.453	0.325	0.204	79,752
High school fixed effects sample	0.147	0.640	0.493	0.379	0.232	56,068
High school fixed effects	0.147	0.616	0.469	0.365	0.218	56,068
Asian American						
In-State						
No high school fixed effects	0.536	0.684	0.148	0.628	0.092	6,017
High school fixed effects sample	0.554	0.706	0.152	0.650	0.096	5,648
High school fixed effects	0.554	0.703	0.149	0.652	0.098	5,648
HS and Census tract fixed effects	0.553	0.716	0.163	0.653	0.101	4,266
Out-of-State						
No high school fixed effects	0.167	0.630	0.464	0.388	0.222	16,202
High school fixed effects sample	0.192	0.683	0.491	0.433	0.241	12,197
High school fixed effects	0.192	0.664	0.472	0.420	0.227	12,197
White						
In-State						
No high school fixed effects	0.509	0.682	0.173	0.615	0.106	37,094
High school fixed effects sample	0.522	0.697	0.175	0.630	0.109	35,173
High school fixed effects	0.522	0.693	0.172	0.633	0.111	35,173
HS and Census tract fixed effects	0.532	0.722	0.191	0.648	0.116	26,005
Out-of-State						
No high school fixed effects	0.109	0.560	0.450	0.310	0.200	63,550
High school fixed effects sample	0.134	0.628	0.493	0.364	0.230	43,871
High school fixed effects	0.134	0.603	0.468	0.350	0.216	43,871

Table 4.4.R: In-State Admissions Probabilities under Counterfactual Regimes

	Number of admits	Share of admits
Data		
White	18,865	68.8%
African American	2,374	8.7%
Hispanic	1,470	5.4%
Asian American	3,223	11.8%
Total	27,422	
No racial preferences		
White	19,889	72.5%
African American	1,532	5.6%
Hispanic	1,212	4.4%
Asian American	3,370	12.3%
No legacy preferences		
White	18,824	68.6%
African American	2,387	8.7%
Hispanic	1,480	5.4%
Asian American	3,237	11.8%
No racial or legacy preferences		
White	19,853	72.4%
African American	1,544	5.6%
Hispanic	1,222	4.5%
Asian American	3,385	12.3%

Table 4.5.R: Out-of-State Admissions Probabilities under Counterfactual Regimes

	Number of admits	Share of admits
Data		
White	6,954	48.7%
African American	1,605	11.2%
Hispanic	1,821	12.8%
Asian American	2,698	18.9%
Total	14,281	
No racial preferences		
White	8,878	62.2%
African American	208	1.5%
Hispanic	738	5.2%
Asian American	3,260	22.8%
No legacy preferences		
White	6,677	46.8%
African American	1,657	11.6%
Hispanic	1,879	13.2%
Asian American	2,841	19.9%
No racial or legacy preferences		
White	8,672	60.7%
African American	209	1.5%
Hispanic	764	5.3%
Asian American	3,422	24.0%

Table 5.1.R: Where African American and Hispanic Applicants and Admits Fall on the Asian American and White Admissions Index Distribution

	Median African American		Median Hispanic		
	Percentile of Applicant Dist	Percentile of Admit Dist	Percentile of Applicant Dist	Percentile of Admit Dist	Observations
Asian American					
In-State					
No high school fixed effects	18%	8%	30%	20%	6,017
High school fixed effects sample	17%	8%	29%	20%	5,648
High school fixed effects	17%	9%	29%	19%	5,648
HS and Census tract fixed effects	16%	7%	30%	19%	4,266
Out-of-State					
No high school fixed effects	8%	1%	29%	7%	16,202
High school fixed effects sample	8%	1%	29%	7%	12,197
High school fixed effects	10%	<1%	32%	7%	12,197
White					
In-State					
No high school fixed effects	16%	10%	30%	24%	37,094
High school fixed effects sample	17%	10%	30%	24%	35,173
High school fixed effects	17%	11%	31%	24%	35,173
HS and Census tract fixed effects	16%	8%	30%	22%	26,005
Out-of-State					
No high school fixed effects	12%	2%	36%	10%	63,550
High school fixed effects sample	10%	2%	36%	10%	43,871
High school fixed effects	14%	1%	39%	8%	43,871

Appendix B

(Tables included in Appendix B are labeled to correspond with the tables in Appendix A of my opening report)

Table A.4.1.R: Logit Estimates of In-State Admissions, 2016-2021

Variable	spec1	spec2	spec3	spec4	spec5	spec6	spec7
African American	-0.589	1.851	2.863	3.542	3.599	3.986	4.729
	(0.029)	(0.057)	(0.073)	(0.119)	(0.12)	(0.14)	(0.18)
Hispanic	-0.131	1.240	1.771	1.993	1.997	2.313	2.605
	(0.038)	(0.070)	(0.09)	(0.15)	(0.15)	(0.16)	(0.21)
Asian American	0.235	-0.133	-0.011	0.148	0.167	0.167	0.176
	(0.029)	(0.057)	(0.07)	(0.10)	(0.11)	(0.12)	(0.14)
female	0.104	0.198	0.035	0.112	0.124	0.177	0.172
	(0.018)	-(0.031)	(0.04)	(0.05)	(0.05)	(0.05)	(0.07)
FGC	-0.304	0.647	0.926	1.168	1.174	1.142	1.298
	(0.024)	(0.04)	(0.05)	(0.06)	(0.07)	(0.07)	(0.09)
regular admission	-0.981	-0.571	-0.503	-0.512	-0.499	-0.604	-0.353
	(0.020)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)
alum	0.193	0.380	0.447	0.467	0.480	0.351	0.426
	(0.025)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(0.07)
waiver	-0.083	0.359	0.277	0.349	0.355	0.165	0.200
	(0.03)	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)	(0.10)
faculty child	0.195	0.502	0.760	0.762	0.754	0.333	0.608
	(0.07)	(0.12)	(0.15)	(0.15)	(0.15)	(0.17)	(0.21)
female * race							
African American				-0.469	-0.516	-0.628	-0.624
				(0.12)	(0.13)	(0.14)	(0.18)
Hispanic				-0.166	-0.109	-0.156	-0.201
				(0.15)	(0.16)	(0.17)	(0.22)
Asian American				-0.247	-0.274	-0.357	-0.283
				(0.12)	(0.12)	(0.13)	(0.16)
FGC * race							
African American				-1.027	-0.985	-1.088	-1.434
				(0.12)	(0.13)	(0.14)	(0.19)
Hispanic				-0.392	-0.343	-0.437	-0.456
				(0.16)	(0.17)	(0.18)	(0.23)
Asian American				-0.148	-0.148	-0.001	-0.125
				(0.14)	(0.15)	(0.16)	(0.20)
Academic variables		X	X	X	X	X	X
Ratings variables			X	X	X	X	X
Heterogeneity variables				X	X	X	X
HS fixed effects sample					X	X	X
HS fixed effects model						X	X
Census tract fixed effects							X
N	57,225	57,225	57,225	57,225	53,504	53,504	38,870
Pseudo R-squared	0.0564	0.588	0.725	0.727	0.724	0.754	0.771

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlsx, UNC0379829.xlsx, UNC0379830.xlsx, UNC0379831.xlsx, UNC0379832.xlsx, UNC0379833.xlsx, UNC0379838.xlsx.

Notes: The academic variables include SAT math and verbal scores (entered as z-scores), percentile, rank, GPA, percentile*rank, GPA*race, rank within the group of applicants from the same high school, indicator for minimum requirements not being met, plus controls for missing variables. The ratings variables include program score, performance, activities, essay scores, and personal quality. The heterogeneity variables include intended major, missing intended major, female*race, and FGC*race.

Table A.4.2.R: Logit Estimates of Out-of-State Admissions, 2016-2021

Variable	spec1	spec2	spec3	spec4	spec5	spec6
African American	0.866	4.766	5.934	6.162	7.190	6.284
	(0.033)	(0.077)	(0.095)	(0.125)	(0.188)	(0.156)
Hispanic	0.980	2.484	3.054	3.000	3.553	3.129
	(0.031)	(0.071)	(0.083)	(0.104)	(0.147)	(0.124)
Asian American	0.781	0.196	0.090	0.077	0.201	0.151
	(0.026)	(0.055)	(0.065)	(0.079)	(0.109)	(0.091)
female	-0.157	0.333	0.032	-0.075	-0.094	-0.065
	(0.019)	(0.025)	(0.030)	(0.040)	(0.053)	(0.045)
FGC	-0.172	0.912	1.367	1.889	2.453	2.026
	(0.033)	(0.044)	(0.052)	(0.075)	(0.111)	(0.093)
regular admission	-0.846	-0.727	-0.809	-0.828	-0.967	-0.820
	(0.020)	(0.025)	(0.030)	(0.030)	(0.041)	(0.034)
alum	1.866	3.412	4.741	4.769	5.649	5.001
	(0.037)	(0.055)	(0.071)	(0.072)	(0.098)	(0.082)
waiver	-0.135	0.360	0.259	0.349	0.156	0.270
	(0.039)	(0.051)	(0.060)	(0.061)	(0.089)	(0.077)
female * race						
African American				0.081	0.074	0.112
				(0.107)	(0.150)	(0.130)
Hispanic				0.357	0.298	0.314
				(0.094)	(0.126)	(0.107)
Asian American				0.107	0.045	0.065
				(0.075)	(0.096)	(0.083)
FGC * race						
African American				-1.343	-1.518	-1.215
				(0.136)	(0.204)	(0.176)
Hispanic				-0.986	-1.193	-0.919
				(0.136)	(0.195)	(0.170)
Asian American				-0.554	-0.685	-0.542
				(0.130)	(0.176)	(0.152)
Academic variables		X	X	X	X	X
Ratings variables			X	X	X	X
Heterogeneity variables				X	X	X
HS fixed effects sample					X	X
HS fixed effects model						X
N	105,623	105,623	105,137	105,116	72,397	72,397
Pseudo R-squared	0.0727	0.42	0.586	0.588	0.645	0.584

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlsx, UNC0379829.xlsx.

Notes: The academic variables include SAT math and verbal scores (entered as z-scores), percentile, rank, GPA, percentile*rank, GPA*race, rank within the group of applicants from the same high school, indicator for minimum requirements not being met, plus controls for missing variables. The ratings variables include program score, performance, activities, essay scores, and personal quality. The heterogeneity variables include intended major, missing intended major, female*race, and FGC*race.

Table A.4.3.R: In-State Admissions Probabilities under Counterfactual Regimes

	Number of admits							Share of admits						
	2016	2017	2018	2019	2020	2021	Total	2016	2017	2018	2019	2020	2021	Total
Data														
White	3,021	3,003	3,134	3,043	3,308	3,356	18,865	71.8%	68.8%	69.3%	69.3%	67.4%	66.8%	68.8%
African American	383	373	405	373	429	411	2,374	9.1%	8.5%	9.0%	8.5%	8.7%	8.2%	8.7%
Hispanic	196	227	249	240	261	297	1,470	4.7%	5.2%	5.5%	5.5%	5.3%	5.9%	5.4%
Asian American	464	469	545	487	603	655	3,223	11.0%	10.7%	12.0%	11.1%	12.3%	13.0%	11.8%
Total	4,207	4,365	4,524	4,394	4,910	5,022	27,422							
No racial preferences														
White	3,237	3,166	3,331	3,200	3,458	3,497	19,889	76.9%	72.5%	73.6%	72.8%	70.4%	69.6%	72.5%
African American	221	244	247	244	274	302	1,532	5.3%	5.6%	5.5%	5.6%	5.6%	6.0%	5.6%
Hispanic	155	174	196	203	231	253	1,212	3.7%	4.0%	4.3%	4.6%	4.7%	5.0%	4.4%
Asian American	476	483	581	506	648	676	3,370	11.3%	11.1%	12.8%	11.5%	13.2%	13.5%	12.3%
No legacy preferences														
White	3,016	2,992	3,129	3,038	3,299	3,350	18,824	71.7%	68.5%	69.2%	69.1%	67.2%	66.7%	68.6%
African American	385	377	407	374	432	412	2,387	9.2%	8.6%	9.0%	8.5%	8.8%	8.2%	8.7%
Hispanic	197	229	250	241	264	299	1,480	4.7%	5.2%	5.5%	5.5%	5.4%	6.0%	5.4%
Asian American	465	473	547	489	606	657	3,237	11.1%	10.8%	12.1%	11.1%	12.3%	13.1%	11.8%
No racial or legacy preferences														
White	3,232	3,157	3,327	3,195	3,450	3,492	19,853	76.8%	72.3%	73.5%	72.7%	70.3%	69.5%	72.4%
African American	223	247	249	245	277	303	1,544	5.3%	5.7%	5.5%	5.6%	5.6%	6.0%	5.6%
Hispanic	156	176	197	204	234	255	1,222	3.7%	4.0%	4.4%	4.6%	4.8%	5.1%	4.5%
Asian American	477	487	583	508	652	678	3,385	11.3%	11.2%	12.9%	11.6%	13.3%	13.5%	12.3%

Table A.4.4.R: Out-of-State Admissions Probabilities under Counterfactual Regimes

	Number of admits							Share of admits						
	2016	2017	2018	2019	2020	2021	Total	2016	2017	2018	2019	2020	2021	Total
Data														
White	999	1,020	1,262	1,391	1,219	1,063	6,954	53.0%	50.9%	49.1%	47.7%	48.2%	44.7%	48.7%
African American	249	227	326	305	227	271	1,605	13.2%	11.3%	12.7%	10.5%	9.0%	11.4%	11.2%
Hispanic	242	239	310	361	339	330	1,821	12.8%	11.9%	12.1%	12.4%	13.4%	13.9%	12.8%
Asian American	293	323	478	582	485	537	2,698	15.6%	16.1%	18.6%	20.0%	19.2%	22.6%	18.9%
Total	1,884	2,004	2,570	2,915	2,529	2,379	14,281							
No racial preferences														
White	1,351	1,297	1,651	1,737	1,515	1,327	8,878	71.7%	64.7%	64.2%	59.6%	59.9%	55.8%	62.2%
African American	17	26	47	49	23	46	208	0.9%	1.3%	1.8%	1.7%	0.9%	1.9%	1.5%
Hispanic	61	104	129	149	134	161	738	3.2%	5.2%	5.0%	5.1%	5.3%	6.8%	5.2%
Asian American	389	341	567	703	589	671	3,260	20.6%	17.0%	22.1%	24.1%	23.3%	28.2%	22.8%
No legacy preferences														
White	967	982	1,214	1,350	1,166	998	6,677	51.3%	49.0%	47.2%	46.3%	46.1%	42.0%	46.8%
African American	257	234	337	312	235	282	1,657	13.6%	11.7%	13.1%	10.7%	9.3%	11.9%	11.6%
Hispanic	246	249	319	371	351	343	1,879	13.1%	12.4%	12.4%	12.7%	13.9%	14.4%	13.2%
Asian American	310	342	503	605	511	570	2,841	16.5%	17.1%	19.6%	20.8%	20.2%	24.0%	19.9%
No racial or legacy preferences														
White	1,329	1,269	1,618	1,707	1,474	1,275	8,672	70.5%	63.3%	63.0%	58.6%	58.3%	53.6%	60.7%
African American	18	27	48	49	21	46	209	1.0%	1.3%	1.9%	1.7%	0.8%	1.9%	1.5%
Hispanic	63	109	133	152	140	167	764	3.3%	5.4%	5.2%	5.2%	5.5%	7.0%	5.3%
Asian American	409	361	594	729	620	709	3,422	21.7%	18.0%	23.1%	25.0%	24.5%	29.8%	24.0%

Table A.4.5.R: Counterfactual Analysis Year-by-Year Logit Estimates of In-State Admissions, 2016-2021

Variable	2016	2017	2018	2019	2020	2021
African American	4.840	3.559	4.580	3.975	3.707	2.814
	(0.321)	(0.294)	(0.326)	(0.352)	(0.284)	(0.299)
Hispanic	2.607	3.015	3.202	1.940	1.223	1.761
	(0.401)	(0.422)	(0.406)	(0.416)	(0.320)	(0.356)
Asian American	0.440	0.326	-0.019	0.070	-0.134	0.327
	(0.272)	(0.287)	(0.269)	(0.297)	(0.244)	(0.260)
female	0.125	0.020	0.156	-0.013	0.253	0.050
	(0.118)	(0.122)	(0.123)	(0.124)	(0.112)	(0.112)
FGC	1.240	1.636	1.322	1.372	0.828	1.260
	(0.153)	(0.160)	(0.167)	(0.167)	(0.157)	(0.158)
regular admission	-0.388	-0.283	-0.315	-0.449	0.047	-2.065
	(0.100)	(0.104)	(0.113)	(0.115)	(0.104)	(0.126)
alum	0.448	0.772	0.390	0.407	0.600	0.382
	(0.131)	(0.138)	(0.135)	(0.133)	(0.123)	(0.126)
waiver	0.279	0.317	0.451	0.336	0.325	0.592
	(0.177)	(0.177)	(0.168)	(0.168)	(0.148)	(0.146)
Faculty Child	0.683	0.861	0.681	0.720	1.388	0.469
	(0.477)	(0.486)	(0.371)	(0.362)	(0.353)	(0.333)
female * race						
African American	-0.640	-0.268	-0.567	-0.533	-0.613	-0.447
	(0.316)	(0.309)	(0.324)	(0.351)	(0.287)	(0.308)
Hispanic	-0.800*	0.150	-0.276	-0.439	-0.312	0.105
	(0.414)	(0.434)	(0.424)	(0.434)	(0.348)	(0.346)
Asian American	0.055	-0.151	-0.189	0.031	-0.532	-0.495
	(0.329)	(0.329)	(0.314)	(0.340)	(0.290)	(0.296)
FGC * race						
African American	-1.563	-1.276	-1.446	-0.843	-1.043	-0.805
	(0.317)	(0.319)	(0.334)	(0.349)	(0.303)	(0.319)
Hispanic	0.001	-1.492	-1.060	0.642	0.218	-0.753
	(0.435)	(0.444)	(0.448)	(0.457)	(0.371)	(0.365)
Asian American	-0.502	-0.220	-0.175	-0.205	0.142	-0.156
	(0.375)	(0.372)	(0.361)	(0.410)	(0.350)	(0.358)
N	8,625	8,455	8,491	8,708	10,337	10,358
Pseudo R-squared	0.728	0.729	0.736	0.750	0.731	0.738

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlsx, UNC0379829.xlsx

Notes: The academic variables include SAT math and verbal scores (entered as z-scores), percentile, rank, GPA, percentile*rank, GPA*race, rank within the group of applicants from the same high school, indicator for minimum requirements not being met, plus controls for missing variables. The ratings variables include program score, performance, activities, essay scores, and personal quality. The heterogeneity variables include intended major, missing intended major, female*race, and FGC*race.

Table A.4.6.R: Counterfactual Analysis Year-by-Year Logit Estimates of Out-of-State Admissions, 2016-2021

Variable	2016	2017	2018	2019	2020	2021
African American	7.322	5.356	6.397	7.064	6.616	6.023
	(0.329)	(0.343)	(0.290)	(0.325)	(0.326)	(0.315)
Hispanic	4.104	2.782	2.621	3.495	3.291	2.734
	(0.304)	(0.265)	(0.252)	(0.266)	(0.266)	(0.252)
Asian American	-0.002	0.440	0.231	-0.078	-0.090	-0.006
	(0.214)	(0.220)	(0.192)	(0.197)	(0.198)	(0.191)
female	0.046	-0.235	-0.035	0.068	-0.125	-0.165
	(0.103)	(0.104)	(0.096)	(0.097)	(0.101)	(0.104)
FGC	1.883	2.076	1.315	2.357	1.379	2.719
	(0.193)	(0.189)	(0.189)	(0.178)	(0.203)	(0.186)
regular admission	-0.893	-1.005	-0.663	-0.222	-1.663	-0.719
	(0.079)	(0.078)	(0.072)	(0.071)	(0.082)	(0.079)
alum	4.767	4.497	4.548	4.973	5.114	5.994
	(0.191)	(0.173)	(0.178)	(0.188)	(0.188)	(0.192)
waiver	-0.041	0.282	0.611	0.572	0.122	0.307
	(0.181)	(0.185)	(0.138)	(0.150)	(0.157)	(0.144)
female * race						
African American	0.325	0.951	0.089	-0.332	-0.227	-0.410
	(0.272)	(0.293)	(0.253)	(0.277)	(0.279)	(0.268)
Hispanic	0.534**	0.195	0.765	0.328	0.196	0.228
	(0.273)	(0.249)	(0.231)	(0.235)	(0.231)	(0.227)
Asian American	0.172	0.154	-0.052	0.189	0.174	-0.048
	(0.208)	(0.208)	(0.183)	(0.184)	(0.188)	(0.180)
FGC * race						
African American	-1.337	-1.505	-0.801	-1.848	-1.167	-1.918
	(0.348)	(0.343)	(0.333)	(0.354)	(0.374)	(0.334)
Hispanic	-1.061	-0.863	-0.031	-1.058	-1.162	-2.024
	(0.364)	(0.355)	(0.340)	(0.315)	(0.364)	(0.350)
Asian American	-0.844	-0.660	-0.150	-0.681	-0.241	-0.750
	(0.358)	(0.351)	(0.322)	(0.309)	(0.347)	(0.303)
N	15,045	15,990	16,311	16,303	18,380	20,745
Pseudo R-squared	0.578	0.582	0.586	0.617	0.623	0.623

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlsx, UNC0379829.xlsx

Notes: The academic variables include SAT math and verbal scores (entered as z-scores), percentile, rank, GPA, percentile*rank, GPA*race, rank within the group of applicants from the same high school, indicator for minimum requirements not being met, plus controls for missing variables. The ratings variables include program score, performance, activities, essay scores, and personal quality. The heterogeneity variables include intended major, missing intended major, female*race, and FGC*race.

Table A.5.1.R: Ratings Ordered Logit Estimates for In-State Applicants, 2016-2021

	Program						Performance						Activity					
Variable	spec1	spec2	spec3	spec4	spec5	spec6	spec1	spec2	spec3	spec4	spec5	spec6	spec1	spec2	spec3	spec4	spec5	spec6
African American	-0.487	0.362	0.251	0.232	0.224	0.098	-1.143	-0.366	-0.276	-0.172	-0.163	0.095	-0.557	-0.129	-0.194	-0.123	-0.128	-0.046
	(0.025)	(0.028)	(0.028)	(0.046)	(0.047)	(0.051)	(0.024)	(0.027)	(0.027)	(0.045)	(0.046)	(0.048)	(0.025)	(0.028)	(0.028)	(0.046)	(0.047)	(0.048)
Hispanic	-0.065	0.322	0.217	0.006	-0.016	0.116	-0.614	-0.303	-0.239	-0.122	-0.124	-0.024	-0.367	-0.160	-0.289	-0.176	-0.179	-0.168
	(0.032)	(0.036)	(0.036)	(0.060)	(0.061)	(0.065)	(0.032)	(0.035)	(0.035)	(0.061)	(0.062)	(0.063)	(0.033)	(0.036)	(0.036)	(0.062)	(0.063)	(0.063)
Asian American	0.846	0.710	0.522	0.549	0.546	0.786	-0.274	-0.706	-0.497	-0.592	-0.605	-0.359	-0.105	-0.275	-0.329	-0.198	-0.208	-0.212
	(0.025)	(0.029)	(0.029)	(0.045)	(0.045)	(0.048)	(0.025)	(0.029)	(0.029)	(0.043)	(0.044)	(0.046)	(0.026)	(0.029)	(0.029)	(0.044)	(0.045)	(0.046)
Female	-0.029	-0.073	-0.016	-0.016	-0.033	0.032	0.412	0.244	0.215	0.230	0.213	0.294	0.130	0.151	0.099	0.147	0.147	0.150
	(0.015)	(0.016)	(0.017)	(0.020)	(0.020)	(0.021)	(0.015)	(0.016)	(0.017)	(0.020)	(0.021)	(0.021)	(0.016)	(0.017)	(0.017)	(0.021)	(0.021)	(0.021)
First Generation	-0.304	-0.022	0.078	0.056	0.045	-0.047	-0.147	0.196	0.210	0.230	0.228	-0.042	-0.543	-0.384	-0.382	-0.347	-0.353	-0.365
	(0.020)	(0.021)	(0.021)	(0.027)	(0.028)	(0.030)	(0.020)	(0.021)	(0.021)	(0.027)	(0.028)	(0.029)	(0.021)	(0.021)	(0.021)	(0.028)	(0.028)	(0.029)
Regular Admission	-0.615	-0.116	-0.143	-0.147	-0.150	-0.223	-0.776	-0.059	-0.096	-0.098	-0.104	-0.108	-0.454	-0.202	-0.185	-0.178	-0.177	-0.181
	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)	(0.017)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.017)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)
Legacy	0.059	0.088	0.046	0.050	0.049	-0.041*	-0.052	-0.081	-0.063	-0.055	-0.056	-0.068	0.225	0.217	0.192	0.186	0.183	0.136
	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.023)	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)
Waiver	0.066	0.286	0.334	0.322	0.334	0.020	-0.012	0.088	0.199	0.209	0.231	0.170	-0.203	-0.114	-0.202	-0.195	-0.189	-0.166
	(0.026)	(0.026)	(0.027)	(0.027)	(0.028)	(0.031)	(0.025)	(0.026)	(0.026)	(0.026)	(0.027)	(0.029)	(0.026)	(0.026)	(0.026)	(0.027)	(0.027)	(0.029)
Faculty Child	-0.207	-0.476	-0.508	-0.501	-0.490	-0.132	-0.130	0.051	-0.143	-0.141	-0.131	-0.033	-0.007	0.046	0.013	0.008	0.015	-0.034
	(0.056)	(0.057)	(0.058)	(0.058)	(0.058)	(0.064)	(0.058)	(0.059)	(0.060)	(0.060)	(0.060)	(0.065)	(0.060)	(0.061)	(0.061)	(0.061)	(0.061)	(0.065)
female * race																		
African American				0.081	0.091	-0.014				-0.093	-0.110	-0.127				-0.157	-0.146	-0.125
				(0.049)	(0.051)	(0.054)				(0.048)	(0.050)	(0.051)				(0.049)	(0.050)	(0.051)
Hispanic				0.089	0.092	0.045				-0.110	-0.126	-0.114				-0.124	-0.137	-0.132
				(0.066)	(0.067)	(0.071)				(0.065)	(0.067)	(0.068)				(0.067)	(0.068)	(0.068)
Asian American				-0.059	-0.050	-0.131				0.072	0.081	0.041				-0.045	-0.035	-0.036
				(0.052)	(0.053)	(0.056)				(0.051)	(0.052)	(0.053)				(0.052)	(0.052)	(0.053)
FGC * race																		
African American				-0.104	-0.083	-0.081				-0.149	-0.125	0.079				0.095	0.086	0.095
				(0.052)	(0.053)	(0.057)				(0.050)	(0.052)	(0.054)				(0.051)	(0.052)	(0.053)
Hispanic				0.335	0.364	0.282				-0.133	-0.112	0.102				-0.061	-0.048	0.015
				(0.069)	(0.071)	(0.075)				(0.068)	(0.069)	(0.071)				(0.070)	(0.071)	(0.072)
Asian American				0.023	0.047	-0.085				0.134	0.139	0.289				-0.290	-0.276	-0.218
				(0.062)	(0.063)	(0.067)				(0.061)	(0.062)	(0.064)				(0.062)	(0.063)	(0.064)
Academic variables		X	X	X	X	X		X	X	X	X	X		X	X	X	X	X
Ratings variables			X	X	X	X			X	X	X	X			X	X	X	X
Heterogeneity variables				X	X	X				X	X	X				X	X	X
HS fixed effects sample					X	X					X	X					X	X
HS fixed effects model						X						X						X
N	57,225	57,225	57,225	57,225	55,294	55,294	57,225	57,225	57,225	57,225	55,294	55,294	57,225	57,225	57,225	57,225	55,294	55,294
Pseudo R-squared	0.0188	0.0997	0.136	0.137	0.142	0.252	0.0282	0.23	0.265	0.266	0.274	0.329	0.0234	0.0497	0.0807	0.0826	0.083	0.0954

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlxs, UNC0379829.xlxs.

Notes: The academic variables include SAT math and verbal scores (entered as z-scores), percentile, rank, GPA, percentile*rank, GPA*race, rank within the group of applicants from the same high school, indicator for minimum requirements not being met, plus controls for missing variables. The ratings variables include program score, performance, activities, essay scores, and personal quality. The heterogeneity variables include intended major, missing intended major, female*race, and FGC*race.

Table A.5.1.R (continued): Ratings Ordered Logit Estimates for In-State Applicants, 2016-2021

	Essay						Personal Quality					
Variable	spec1	spec2	spec3	spec4	spec5	spec6	spec1	spec2	spec3	spec4	spec5	spec6
African American	-0.710	-0.032	-0.091	-0.045	-0.042	-0.012	0.009	0.409	0.477	0.628	0.637	0.656
	(0.034)	(0.038)	(0.039)	(0.063)	(0.065)	(0.067)	(0.034)	(0.039)	(0.041)	(0.067)	(0.069)	(0.071)
Hispanic	-0.085	0.217	0.145	-0.005	-0.016	-0.036	0.278	0.510	0.539	0.447	0.441	0.471
	(0.045)	(0.050)	(0.052)	(0.090)	(0.091)	(0.092)	(0.043)	(0.048)	(0.050)	(0.088)	(0.089)	(0.090)
Asian American	0.205	0.090	0.147	-0.043	-0.046	-0.093	0.121	0.031	0.081	0.052	0.079	0.048
	(0.036)	(0.040)	(0.042)	(0.064)	(0.064)	(0.067)	(0.035)	(0.040)	(0.043)	(0.066)	(0.066)	(0.069)
Female	0.217	0.327	0.276	0.202	0.201	0.147	0.109	0.123	-0.002	0.016	0.018	0.002
	(0.022)	(0.023)	(0.024)	(0.030)	(0.031)	(0.031)	(0.022)	(0.023)	(0.025)	(0.031)	(0.031)	(0.032)
First Generation	-0.768	-0.386	-0.351	-0.352	-0.333	-0.199	-0.062	0.172	0.363	0.321	0.328	0.329
	(0.029)	(0.030)	(0.031)	(0.040)	(0.041)	(0.043)	(0.029)	(0.030)	(0.031)	(0.042)	(0.042)	(0.044)
Regular Admission	-0.250	0.002	0.085	0.073	0.075	0.079	-0.299	-0.112	-0.012	-0.016	-0.019	-0.014
	(0.024)	(0.025)	(0.026)	(0.026)	(0.027)	(0.027)	(0.025)	(0.026)	(0.027)	(0.027)	(0.028)	(0.029)
Legacy	0.178	0.133	0.068	0.058	0.058	0.029	0.119	0.088	0.018	0.014	0.013	-0.008
	(0.030)	(0.031)	(0.032)	(0.032)	(0.033)	(0.034)	(0.030)	(0.030)	(0.032)	(0.032)	(0.033)	(0.034)
Waiver	-0.387	-0.130	-0.210	-0.216	-0.227	-0.145	0.394	0.532	0.611	0.603	0.591	0.577
	(0.035)	(0.035)	(0.036)	(0.037)	(0.038)	(0.040)	(0.034)	(0.035)	(0.036)	(0.037)	(0.038)	(0.040)
Faculty Child	0.367	0.214	0.173	0.152	0.152	0.024	0.203	0.182	0.138	0.138	0.129	0.028
	(0.080)	(0.081)	(0.084)	(0.084)	(0.085)	(0.090)	(0.078)	(0.079)	(0.084)	(0.084)	(0.085)	(0.091)
female * race												
African American				-0.023	-0.024	-0.040				-0.205	-0.197	-0.209
				(0.067)	(0.068)	(0.070)				(0.070)	(0.072)	(0.073)
Hispanic				0.212	0.222	0.229				-0.011	-0.007	-0.011
				(0.094)	(0.096)	(0.097)				(0.092)	(0.094)	(0.095)
Asian American				0.369	0.378	0.383				0.004	-0.020	0.004
				(0.073)	(0.074)	(0.075)				(0.075)	(0.076)	(0.077)
FGC * race												
African American				-0.025	-0.011	-0.108				0.011	-0.005	-0.013
				(0.070)	(0.072)	(0.073)				(0.074)	(0.076)	(0.078)
Hispanic				0.064	0.075	-0.020				0.214	0.202	0.172
				(0.098)	(0.100)	(0.102)				(0.096)	(0.098)	(0.100)
Asian American				-0.029	-0.049	-0.125				0.120	0.099	0.125
				(0.087)	(0.089)	(0.091)				(0.090)	(0.091)	(0.094)
Academic variables		X	X	X	X	X		X	X	X	X	X
Ratings variables			X	X	X	X			X	X	X	X
Heterogeneity variables				X	X	X				X	X	X
HS fixed effects sample					X	X					X	X
HS fixed effects model						X						X
N	57,225	57,225	57,225	57,225	55,294	55,294	57,225	57,225	57,225	57,225	55,294	55,294
Pseudo R-squared	0.042	0.11	0.179	0.181	0.18	0.195	0.0314	0.0597	0.152	0.152	0.152	0.163

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlxs, UNC0379829.xlxs.

Notes: The academic variables include SAT math and verbal scores (entered as z-scores), percentile, rank, GPA, percentile*rank, GPA*race, rank within the group of applicants from the same high school, indicator for minimum requirements not being met, plus controls for missing variables. The ratings variables include program score, performance, activities, essay scores, and personal quality. The heterogeneity variables include intended major, missing intended major, female*race, and FGC*race.

Table A.5.2.R: Ratings Ordered Logit Estimates for Out-of-State Applicants, 2016-2021

Variable	Program spec1	Program spec2	Program spec3	Program spec4	Program spec5	Program spec6
African American	-0.536	0.594	0.568	0.438	0.423	0.247
	(0.021)	(0.031)	(0.031)	(0.044)	(0.060)	(0.067)
Hispanic	0.261	0.581	0.553	0.375	0.360	0.157
	(0.021)	(0.034)	(0.034)	(0.045)	(0.058)	(0.065)
Asian American	0.831	0.764	0.762	0.801	0.840	0.916
	(0.016)	(0.028)	(0.028)	(0.035)	(0.044)	(0.049)
Female	-0.022	0.077	0.065	0.042	-0.003	0.060
	(0.011)	(0.012)	(0.012)	(0.015)	(0.019)	(0.020)
First Generation	-0.339	-0.019	0.001	-0.053	-0.007	-0.033
	(0.018)	(0.019)	(0.019)	(0.026)	(0.038)	(0.040)
Regular Admission	-0.249	-0.001	0.007	-0.001	-0.009	-0.019
	(0.011)	(0.012)	(0.012)	(0.012)	(0.014)	(0.015)
Legacy	-0.104	0.067	0.039	0.043	0.056	-0.061
	(0.030)	(0.031)	(0.031)	(0.031)	(0.035)	(0.037)
Waiver	0.006	0.166	0.143	0.123	0.236	0.108
	(0.023)	(0.023)	(0.023)	(0.023)	(0.034)	(0.038)
female * race						
African American				0.164	0.196	0.224
				(0.042)	(0.056)	(0.059)
Hispanic				0.169	0.202	0.129
				(0.042)	(0.052)	(0.057)
Asian American				-0.089***	-0.054	-0.061
				0.052	0.053	0.055
FGC * race						
African American				0.047	0.052	-0.015
				(0.050)	(0.073)	(0.078)
Hispanic				0.265	0.329	0.178
				(0.055)	(0.079)	(0.085)
Asian American				0.023	0.058	0.043
				(0.051)	(0.067)	(0.072)
Academic variables		X	X	X	X	X
Ratings variables			X	X	X	X
Heterogeneity variables				X	X	X
HS fixed effects sample					X	X
HS fixed effects model						X
N	105,631	105,631	105,631	105,631	72,357	72,357
Pseudo R-squared	0.0152	0.0854	0.0888	0.0895	0.0952	0.253

Variable	Performance spec1	Performance spec2	Performance spec3	Performance spec4	Performance spec5	Performance spec6
African American	-1.317	-0.206	-0.231	-0.220	-0.110	0.177
	(0.021)	(0.031)	(0.031)	(0.044)	(0.058)	(0.063)
Hispanic	-0.508	-0.420	-0.425	-0.442	-0.524	-0.145
	(0.020)	(0.033)	(0.033)	(0.044)	(0.056)	(0.059)
Asian American	-0.142	-0.500	-0.459	-0.474	-0.502	-0.585
	(0.016)	(0.027)	(0.027)	(0.034)	(0.041)	(0.044)
Female	0.435	0.564	0.537	0.535	0.544	0.580
	(0.011)	(0.012)	(0.012)	(0.016)	(0.019)	(0.020)
First Generation	-0.344	-0.044	-0.032	-0.059	-0.122	-0.095
	(0.018)	(0.019)	(0.019)	(0.026)	(0.037)	(0.039)
Regular Admission	-0.466	-0.191	-0.173	-0.174	-0.203	-0.244
	(0.011)	(0.012)	(0.012)	(0.012)	(0.014)	(0.015)
Legacy	-0.323	-0.140	-0.162	-0.150	-0.177	-0.152
	(0.030)	(0.031)	(0.031)	(0.031)	(0.035)	(0.037)
Waiver	-0.111	-0.075	-0.092	-0.093	-0.129	0.007
	(0.022)	(0.023)	(0.023)	(0.023)	(0.033)	(0.035)
female * race						
African American				-0.069*	-0.061	-0.063
				(0.042)	(0.055)	(0.056)
Hispanic				0.083**	0.083	0.078
				(0.041)	(0.051)	(0.053)
Asian American				-0.043	-0.047	-0.033
				(0.032)	(0.037)	(0.039)
FGC * race						
African American				0.058	0.158	0.204
				(0.049)	(0.072)	(0.075)
Hispanic				-0.118	-0.001	0.042
				(0.053)	(0.075)	(0.078)
Asian American				0.130	0.131	0.132
				(0.050)	(0.066)	(0.069)
Academic variables		X	X	X	X	X
Ratings variables			X	X	X	X
Heterogeneity variables				X	X	X
HS fixed effects sample					X	X
HS fixed effects model						X
N	105,631	105,631	105,631	105,631	72,357	72,357
Pseudo R-squared	0.0245	0.155	0.159	0.16	0.158	0.235

Variable	Activity spec1	Activity spec2	Activity spec3	Activity spec4	Activity spec5	Activity spec6
African American	-0.605	0.003	-0.189	-0.097	-0.022	0.030
	(0.021)	(0.031)	(0.032)	(0.045)	(0.061)	(0.065)
Hispanic	-0.194	-0.028	-0.163	-0.138	-0.193	-0.169
	(0.021)	(0.034)	(0.034)	(0.045)	(0.057)	(0.060)
Asian American	0.140	-0.054	-0.098	-0.004	0.032	0.003
	(0.016)	(0.027)	(0.028)	(0.035)	(0.042)	(0.045)
Female	0.069	0.198	0.088	0.133	0.159	0.160
	(0.011)	(0.012)	(0.012)	(0.016)	(0.019)	(0.020)
First Generation	-0.573	-0.290	-0.321	-0.278	-0.265	-0.281
	(0.019)	(0.019)	(0.019)	(0.027)	(0.038)	(0.039)
Regular Admission	-0.291	-0.124	-0.096	-0.092	-0.108	-0.113
	(0.012)	(0.012)	(0.012)	(0.012)	(0.014)	(0.015)
Legacy	0.046	0.107	0.061	0.062	0.029	0.038
	(0.031)	(0.031)	(0.032)	(0.032)	(0.036)	(0.037)
Waiver	-0.184	-0.040	-0.192	-0.180	-0.231	-0.188
	(0.023)	(0.023)	(0.023)	(0.024)	(0.034)	(0.036)
female * race						
African American				-0.094	-0.088	-0.113
				(0.043)	(0.057)	(0.058)
Hispanic				-0.040	-0.045	-0.045
				(0.042)	(0.052)	(0.054)
Asian American				-0.107	-0.111	-0.110
				(0.033)	(0.039)	(0.040)
FGC * race						
African American				-0.080	-0.138	-0.106
				(0.051)	(0.074)	(0.076)
Hispanic				-0.004	0.010	0.049
				(0.055)	(0.077)	(0.079)
Asian American				-0.121	-0.187	-0.164
				(0.051)	(0.067)	(0.069)
Academic variables		X	X	X	X	X
Ratings variables			X	X	X	X
Heterogeneity variables				X	X	X
HS fixed effects sample					X	X
HS fixed effects model						X
N	105,631	105,631	105,631	105,631	72,357	72,357
Pseudo R-squared	0.0188	0.0483	0.0902	0.0913	0.0906	0.11

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlxs, UNC0379829.xlxs.

Notes: The academic variables include SAT math and verbal scores (entered as z-scores), percentile, rank, GPA, percentile*rank, GPA*race, rank within the group of applicants from the same high school, indicator for minimum requirements not being met, plus controls for missing variables. The ratings variables include program score, performance, activities, essay scores, and personal quality. The heterogeneity variables include intended major, missing intended major, female*race, and FGC*race.

Table A.5.2.R (continued): Ratings Ordered Logit Estimates for Out-of-State Applicants, 2016-2021

	Essay						Personal Quality					
Variable	spec1	spec2	spec3	spec4	spec5	spec6	spec1	spec2	spec3	spec4	spec5	spec6
African American	-0.501	0.305	0.139	0.202	0.361	0.494	0.150	0.792	0.759	0.814	0.934	0.921
	(0.032)	(0.046)	(0.047)	(0.067)	(0.091)	(0.097)	(0.028)	(0.043)	(0.045)	(0.064)	(0.085)	(0.091)
Hispanic	0.003	0.274	0.156	0.060	0.119	0.188	0.295	0.550	0.551	0.577	0.574	0.548
	(0.029)	(0.049)	(0.051)	(0.067)	(0.086)	(0.092)	(0.026)	(0.044)	(0.047)	(0.061)	(0.079)	(0.084)
Asian American	0.438	0.241	0.256	0.162	0.157	0.171	0.199	0.040	-0.055	-0.082	-0.081	-0.011
	(0.021)	(0.037)	(0.039)	(0.049)	(0.060)	(0.065)	(0.021)	(0.037)	(0.040)	(0.049)	(0.060)	(0.065)
Female	0.203	0.388	0.321	0.252	0.252	0.233	0.028	0.149	-0.017	-0.019	-0.003	-0.008
	(0.016)	(0.017)	(0.018)	(0.024)	(0.028)	(0.030)	(0.015)	(0.016)	(0.018)	(0.023)	(0.028)	(0.030)
First Generation	-0.646	-0.149	-0.131	-0.117	-0.032	0.034	-0.137	0.191	0.332	0.292	0.262	0.314
	(0.028)	(0.028)	(0.029)	(0.041)	(0.059)	(0.061)	(0.025)	(0.026)	(0.028)	(0.040)	(0.057)	(0.059)
Regular Admission	-0.291	-0.085	-0.025	-0.034	-0.081	-0.085	-0.258	-0.095	-0.030	-0.036	-0.049	-0.020
	(0.016)	(0.017)	(0.017)	(0.017)	(0.021)	(0.022)	(0.015)	(0.016)	(0.017)	(0.017)	(0.020)	(0.022)
Legacy	0.142	0.175	0.125	0.104	0.136	0.138	0.174	0.216	0.176	0.170	0.215	0.170
	(0.042)	(0.043)	(0.045)	(0.045)	(0.052)	(0.054)	(0.040)	(0.041)	(0.044)	(0.044)	(0.050)	(0.052)
Waiver	-0.145	0.124	0.005	0.006	0.041	0.023	0.419	0.621	0.660	0.649	0.667	0.650
	(0.033)	(0.033)	(0.034)	(0.034)	(0.049)	(0.051)	(0.028)	(0.029)	(0.031)	(0.031)	(0.044)	(0.047)
female * race												
African American				0.027	-0.018	-0.040				-0.054	-0.046	-0.030
				(0.063)	(0.084)	(0.087)				(0.060)	(0.077)	(0.080)
Hispanic				0.134	0.139	0.101				-0.120	-0.126	-0.146
				(0.063)	(0.077)	(0.081)				(0.057)	(0.070)	(0.074)
Asian American				0.144	0.178	0.191				0.026	-0.001	-0.004
				(0.046)	(0.054)	(0.056)				(0.046)	(0.054)	(0.056)
FGC * race												
African American				-0.201	-0.261	-0.252				-0.014	-0.047	-0.065
				(0.074)	(0.112)	(0.116)				(0.071)	(0.103)	(0.108)
Hispanic				0.036	0.028	0.015				0.185	0.057	-0.002
				(0.082)	(0.116)	(0.121)				(0.074)	(0.105)	(0.110)
Asian American				0.144	0.091	0.072				0.109	0.122	0.081
				(0.076)	(0.099)	(0.103)				(0.072)	(0.095)	(0.099)
Academic variables		X	X	X	X	X		X	X	X	X	X
Ratings variables			X	X	X	X			X	X	X	X
Heterogeneity variables				X	X	X				X	X	X
HS fixed effects sample					X	X					X	X
HS fixed effects model						X						X
N	105,631	105,631	105,631	105,631	72,357	72,357	105,631	105,631	105,631	105,631	72,357	72,357
Pseudo R-squared	0.0231	0.097	0.175	0.177	0.167	0.195	0.0339	0.0735	0.185	0.186	0.187	0.213

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlxs, UNC0379829.xlxs.

Notes: The academic variables include SAT math and verbal scores (entered as z-scores), percentile, rank, GPA, percentile*rank, GPA*race, rank within the group of applicants from the same high school, indicator for minimum requirements not being met, plus controls for missing variables. The ratings variables include program score, performance, activities, essay scores, and personal quality. The heterogeneity variables include intended major, missing intended major, female*race, and FGC*race.

Table A.5.3.R: Index values by race for model specification #4 in standard deviation units

	Program	Performance	Activities	Essay	Personal Quality
In-State					
African American	-0.694	-0.757	-0.791	-0.865	-0.377
Hispanic	-0.311	-0.411	-0.472	-0.523	-0.140
Asian American	0.327	-0.042	0.050	-0.079	0.016
Out-of-State					
African American	-1.097	-1.047	-0.721	-0.762	-0.454
Hispanic	-0.240	-0.236	-0.189	-0.240	-0.107
Asian American	0.176	0.088	0.161	0.088	0.186

Appendix C

Table C.1.: Logit Estimates of In-State Admissions, 2016-2021

Variable	Preferred Model	Robustness 1	Robustness 2
African American	3.542	3.248	3.830
	(0.119)	(0.111)	(0.131)
Hispanic	1.993	1.691	2.233
	(0.148)	(0.134)	(0.168)
Asian American	0.148	0.233	0.275
	(0.104)	(0.096)	(0.121)
female	0.112	-0.039	-0.028
	(0.046)	(0.045)	(0.045)
FGC	1.168	1.067	1.097
	(0.063)	(0.062)	(0.062)
regular admission	-0.512	-0.529	-0.524
	(0.042)	(0.042)	(0.042)
alum	0.467	0.475	0.474
	(0.051)	(0.050)	(0.050)
waiver	0.349	0.350	0.337
	(0.063)	(0.062)	(0.062)
fsch	0.762	0.818	0.816
	(0.148)	(0.145)	(0.145)
female * race	-0.469	-0.455	-0.464
African American	(0.121)	(0.119)	(0.120)
	-0.166	-0.125	-0.152
Hispanic	(0.152)	(0.150)	(0.151)
	-0.247	-0.231	-0.234
Asian American	(0.121)	(0.119)	(0.120)
FGC * race	-1.027	-1.017	-1.014
African American	(0.124)	(0.122)	(0.123)
	-0.392	-0.203	-0.330
Hispanic	(0.159)	(0.152)	(0.158)
	-0.148	-0.118	-0.128
Asian American	(0.143)	(0.140)	(0.142)
N	57,225	57,225	57,224
Pseudo R-squared	0.727	0.720	0.722

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlsx, UNC0379829.xlsx, UNC0379830.xlsx, UNC0379831.xlsx, UNC0379832.xlsx, UNC0379833.xlsx, UNC0379838.xlsx

Table C.2.: Logit Estimates of Out-of-State Admissions, 2016-2021

Variable	Preferred Model	Robustness 1	Robustness 2
African American	6.162	5.161	7.570
	(0.125)	(0.099)	(0.162)
Hispanic	3.000	2.774	3.598
	(0.104)	(0.078)	(0.132)
Asian American	0.077	0.142	-0.185
	(0.079)	(0.059)	(0.101)
female	-0.075	-0.281	-0.260
	(0.040)	(0.039)	(0.039)
FGC	1.889	1.800	1.883
	(0.075)	(0.074)	(0.075)
regular admission	-0.828	-0.855	-0.861
	(0.030)	(0.030)	(0.030)
alum	4.769	4.695	4.777
	(0.072)	(0.071)	(0.072)
waiver	0.349	0.295	0.334
	(0.061)	(0.060)	(0.060)
female * race			
African American	0.081	0.148	0.081
	(0.107)	(0.105)	(0.106)
Hispanic	0.357	0.360	0.359
	(0.094)	(0.093)	(0.094)
Asian American	0.107	0.192	0.177
	(0.075)	(0.075)	(0.075)
FGC * race			
African American	-1.343	-1.272	-1.375
	(0.136)	(0.132)	(0.135)
Hispanic	-0.986	-0.953	-0.995
	(0.136)	(0.134)	(0.136)
Asian American	-0.554	-0.558	-0.614
	(0.130)	(0.128)	(0.129)
N	105,116	105,116	105,116
Pseudo R-squared	0.588	0.580	0.587

Sources: MainDataA.csv, MainDataB.csv, MainDataC.csv, MainDataD.csv, UNC0379828.xlsx, UNC0379829.xlsx, UNC0379830.xlsx, UNC0379831.xlsx, UNC0379832.xlsx, UNC0379833.xlsx, UNC0379838.xlsx

Appendix D

1 Expected errors conditional on choices

In Appendix D of my original report, I defined a latent index π_i where i indexes individuals and where

$$\pi_i = X_i\gamma + \varepsilon_i \tag{1}$$

The university accepts individual i if $\pi_i > 0$. In the above equation, X_i represents attributes about candidate i that I observe in the data. The ε_i represents the unobserved characteristic of the individual. To characterize the role of unobservables, I need to be able to calculate the expected value of ε_i conditional on the admission decision.

A mathematically equivalent model–one that leads to same the estimation procedure and model predictions–would be instead to define the payoff the university receives from accepting the applicant and rejecting the applicant respectively as u_{1i} and u_{0i} and where:

$$\begin{aligned} u_{1i} &= X_i\gamma + \epsilon_{1i} \\ u_{0i} &= \epsilon_{0i} \end{aligned}$$

The university admits the applicants when $u_{1i} - u_{0i} > 0$. Note that ε in equation (1) is then identical to $\epsilon_{1i} - \epsilon_{0i}$. I want to recover $E(\varepsilon|y = 1)$ where y indicates admission. This is the same as recovering $E(\epsilon_1 - \epsilon_0|y = 1)$ but this second way is mathematically easier to derive the expectation.

Under this second way of expressing the logit model, ϵ_1 and ϵ_0 are distributed Type 1 extreme value. This error distribution has the following property:

$$E(\epsilon_0) = \gamma = Pr(y = 0)E(\epsilon_0|y = 0) + Pr(y = 1)E(\epsilon_0|y = 1) \tag{2}$$

where γ is Euler's constant.

Rearranging terms yields:

$$E(\epsilon_0|y = 1) = \frac{\gamma - Pr(y = 0)E(\epsilon_0|y = 0)}{Pr(y = 1)} \tag{3}$$

The previous literature has shown that $E(\epsilon_0|y = 0)$ can be expressed as:[1]

$$E(\epsilon_0|y = 0) = \gamma - \ln(Pr(y = 0)) \tag{4}$$

Substituting (4) into (3) yields:

$$\begin{aligned} E(\epsilon_0|y = 1) &= \frac{\gamma - Pr(y = 0)[\gamma - \ln(Pr(y = 0))]}{Pr(y = 1)} \qquad (5) \\ &= \gamma + \frac{Pr(y = 0)\ln(Pr(y = 0))}{Pr(y = 1)} \qquad (6) \end{aligned}$$

Recognizing that:

$$E(\epsilon_1|y = 1) = \gamma - \ln(Pr(y = 1)) \tag{7}$$

[1] See, for example, V.J. Hotz and R.A. Miller "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, Vol. 60, No.3, July 1993., page 504.

we can form $E(\epsilon_1 - \epsilon_0|y = 1)$ using:

$$E(\epsilon_1 - \epsilon_0|y = 1) = -\ln(Pr(y = 1)) - \frac{Pr(y = 0)\ln(Pr(y = 0))}{Pr(y = 1)} \tag{8}$$

The individual probabilities of admission ($Pr(y = 1)$) and rejection ($Pr(y = 0)$) then translate directly into how strong we expect the applicant to be on unobserved characteristics conditional on being admitted. This can then be compared to the estimated admissions preference which I label μ (this is the coefficient on race in the logit model) to see how often the expected unobserved characteristic is bigger than the racial preference.

2 Probability of unobserved draws

The previous section showed how to calculate the expected value of the unobservable characteristic conditional on three pieces of information: (i) the distribution of the unobserved characteristic, (ii) the probability the individual was admitted, and (iii) whether the individual was actually admitted. These three pieces of information can also be used to calculate the probability the unobserved characteristic is bigger than the racial preference for each applicant.

The probability of the unobserved factor being greater than μ is given by one minus the logistic cumulative distribution function.

$$Pr(\epsilon > \mu) = 1 - \frac{1}{1 + \exp(-\mu)} \tag{9}$$

The probability of the unobserved factor being greater than μ conditional on being admitted given observed characteristics x, where these observed characteristics x translate into an admit probability of $Pr(y = 1)$, is given by:

$$Pr(\epsilon > \mu|y = 1, x) = \min\left\{\frac{1 - \frac{1}{1+\exp(-\mu)}}{Pr(y = 1)}, 1\right\} \tag{10}$$

The reason for the min operator is that some individuals would have x's such that it is assured that their unobservable characteristics had to be bigger than μ.